# ISYE 6644 — Simulation
Course Notes (Transcribed from Handwritten Originals)

## Contents

# 1 Bivariate Distributions and Functions of Random Variables

## 1.1 Functions of a single RV (warmup)

To find the pdf of $Y = h(X)$ when $X$ is continuous, set up the CDF of $Y$ in terms of $X$ and then differentiate. **Watch out for the boundary between discrete and continuous setups!**

**Example 1.1** (Mixed case)**.** Let $X$ have pdf $f(x) = |x|$ for $-1 \le x \le 1$, and let $Y = X^2$. Find $f_Y(y)$.

Set up the CDF for the continuous part:

$$F_Y(y) \;=\; P(Y \le y) \;=\; P(X^2 \le y) \;=\; P(-\sqrt{y} \le X \le \sqrt{y}) \;=\; \int_{-\sqrt{y}}^{\sqrt{y}} |x|\,dx \;=\; y.$$

But the result is a mix of discrete and continuous behavior:

$$F_Y(y) \;=\; \begin{cases} 0 & y \le 0, \\ y & 0 < y < 1, \\ 1 & y \ge 1. \end{cases}$$

Differentiating on the interior gives the pdf:

$$f_Y(y) \;=\; \begin{cases} 1 & 0 < y < 1, \\ 0 & \text{otherwise.} \end{cases}$$

## 1.2 Bivariate case: discrete and continuous

**Discrete joint pmf $f(x, y)$.** Requires $0 \le f(x, y) \le 1$ and $\sum_x \sum_y f(x, y) = 1$. For $A \subseteq \mathbb{R}^2$,

$$P\big((X, Y) \in A\big) \;=\; \sum_{(x,y)\in A} f(x, y).$$

**Continuous joint pdf $f(x, y)$.** Requires $f(x, y) \ge 0$ and $\iint_{\mathbb{R}^2} f(x, y)\,dx\,dy = 1$. For $A \subseteq \mathbb{R}^2$,

$$P\big((X, Y) \in A\big) \;=\; \iint_A f(x, y)\,dx\,dy.$$

The interpretation is

$$f(x, y)\,dy\,dx \;\approx\; P\big(x \le X \le x + dx,\; y \le Y \le y + dy\big).$$

**Example 1.2** (Dartboard / Monte Carlo $\pi$)**.** Let $C \equiv \{(x, y) : (x - \frac{1}{2})^2 + (y - \frac{1}{2})^2 \le \frac{1}{4}\}$ be the inscribed unit circle inside the unit square. Then $\text{Area}(C) = \pi/4$, so the joint pdf of a uniform draw $(X, Y)$ from $C$ is

$$f(x, y) \;=\; \begin{cases} 4/\pi & (x, y) \in C, \\ 0 & \text{otherwise.} \end{cases}$$

Throwing darts at the unit square,

$$\frac{\#\text{ in circle}}{\#\text{ in square}} \;\longrightarrow\; \frac{\text{Area}(C)}{\text{Area(Sq)}} \;=\; \frac{\pi}{4},$$

which gives the Monte Carlo estimator $\pi \approx 4 \cdot (\#\text{in circle}/\#\text{in square})$.

## 1.3 Joint and marginal CDFs / PDFs

The joint CDF is $F(x, y) = P(X \leq x, Y \leq y)$. Key properties:

- Non-decreasing in each argument.
- $\lim_{x \to -\infty} F(x, y) = \lim_{y \to -\infty} F(x, y) = 0$.
- $P(Y \leq y) = \lim_{x \to +\infty} F(x, y) = F_Y(y)$, and similarly $F_X(x) = \lim_{y \to +\infty} F(x, y)$.
- $\lim_{x,y \to \infty} F(x, y) = 1$.

Pass from CDF to PDF by partial differentiation; pass from PDF to CDF by integration.

**Example 1.3.** $F(x, y) = 1 - e^{-x} - e^{-y} + e^{-(x+y)}$ for $x, y \geq 0$, and 0 otherwise. Then

$$F_X(x) \;=\; \lim_{y \to \infty} F(x, y) \;=\; 1 - e^{-x}, \quad x \geq 0,$$

and the joint pdf is

$$f(x, y) \;=\; \frac{\partial^2}{\partial x \partial y} F(x, y) \;=\; e^{-(x+y)}, \quad x, y \geq 0.$$

**Marginal distributions.** "Everything but the variable you want."

$$\text{Discrete: } f_X(x) = \sum_y f(x, y), \quad f_Y(y) = \sum_x f(x, y).$$

$$\text{Continuous: } f_X(x) = \int_{\mathbb{R}} f(x, y)\, dy, \quad f_Y(y) = \int_{\mathbb{R}} f(x, y)\, dx.$$

**Example 1.4.** Joint pdf $f(x, y) = \frac{21}{4} x^2 y$ for $x^2 \leq y \leq 1$. The limits matter!

$$f_X(x) \;=\; \int_{x^2}^{1} \tfrac{21}{4} x^2 y\, dy \;=\; \tfrac{21}{8} x^2 (1 - x^4), \quad -1 \leq x \leq 1.$$

$$f_Y(y) \;=\; \int_{-\sqrt{y}}^{\sqrt{y}} \tfrac{21}{4} x^2 y\, dx \;=\; \tfrac{7}{2} y^{5/2}, \quad 0 \leq y \leq 1.$$

## 1.4 Conditional distributions

By analogy with $P(A \mid B) = P(A \cap B)/P(B)$:

$$f(y \mid x) \;=\; \frac{f(x, y)}{f_X(x)}.$$

This holds in both discrete and continuous cases.

**Example 1.5.** Continuing the previous example with $f(x, y) = \frac{21}{4} x^2 y$ on $x^2 \leq y \leq 1$,

$$f(y \mid x) \;=\; \frac{(21/4) x^2 y}{(21/8) x^2 (1 - x^4)} \;=\; \frac{2y}{1 - x^4}, \quad x^2 \leq y \leq 1.$$

**Inverse-style problems.** Given $f(y \mid x)$ and $f_X(x)$, find $f_Y(y)$:

1. Compute the joint pdf $f(x,y) = f(y \mid x) \cdot f_X(x)$.
2. Marginalize: $f_Y(y) = \int_{\mathbb{R}} f(x,y)\,dx$.

**Example 1.6.** $f_X(x) = 2x$ for $0 < x \le 1$, and $f(y \mid x) \sim \mathrm{Unif}(0,x)$. So $f(y \mid x) = 1/x$ on $0 < y < x$, and

$$f(x,y) \;=\; \frac{1}{x} \cdot 2x \;=\; 2, \quad 0 < y < x < 1.$$

Then $f_Y(y) = \int_y^1 2\,dx = 2(1-y)$ for $0 < y \le 1$.

## 1.5 Independence

$X$ and $Y$ are independent iff $f(x,y) = f_X(x) f_Y(y)$. Equivalently, the joint factors as $f(x,y) = a(x)\,b(y)$ for some functions $a, b$ *with matching support* — limits that couple $x$ and $y$ break independence.

**Example 1.7** (Limits matter)**.** $f(x,y) = c$ on $x^2 \le y \le 1$. Even though the constant pdf "looks" separable, the support couples $x$ and $y$, so $X$ and $Y$ are *not* independent.

**Example 1.8.** $f(x,y) = \dfrac{c}{x+y}$ cannot be written as $a(x)\,b(y)$, so $X$ and $Y$ are not independent.

## 1.6 Expectation, covariance, correlation

Whether or not $X$ and $Y$ are independent,

$$\mathbb{E}[X+Y] = \mathbb{E}[X] + \mathbb{E}[Y], \qquad \mathbb{E}\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} \mathbb{E}[X_i].$$

If $X, Y$ are *independent*,

$$\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y], \qquad \mathrm{Var}(X+Y) = \mathrm{Var}(X) + \mathrm{Var}(Y),$$

and more generally

$$\mathrm{Var}\Big[\sum a_i X_i + b_i\Big] \;=\; \sum a_i^2 \,\mathrm{Var}(X_i).$$

**Covariance.**

$$\mathrm{Cov}(X,Y) \;\equiv\; \sigma_{XY} \;\equiv\; \mathbb{E}\big[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])\big] \;=\; \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y].$$

$\mathrm{Cov}(X,X) = \mathrm{Var}(X)$. Independence implies $\mathrm{Cov}(X,Y) = 0$, but **the converse is false**.

**Example 1.9** (Zero covariance, not independent)**.** $X \sim \mathrm{Unif}(-1,1)$, $Y = X^2$. Then $\mathbb{E}[X] = 0$, $\mathbb{E}[XY] = \mathbb{E}[X^3] = 0$, so $\mathrm{Cov}(X,Y) = 0$ — yet $X$ and $Y$ are perfectly dependent.

**Useful covariance algebra.**

$$\begin{aligned}
\mathrm{Cov}(X+a,\, Y+b) &= \mathrm{Cov}(X,Y),\\
\mathrm{Cov}(aX,\, bY) &= ab\ \mathrm{Cov}(X,Y),\\
\mathrm{Var}(X \pm Y) &= \mathrm{Var}(X) + \mathrm{Var}(Y) \pm 2\ \mathrm{Cov}(X,Y),\\
\mathrm{Cov}(aX + bY,\, cW + dV) &= ac\ \mathrm{Cov}(X,W) + ad\ \mathrm{Cov}(X,V)\\
&\quad + bc\ \mathrm{Cov}(Y,W) + bd\ \mathrm{Cov}(Y,V).
\end{aligned}$$

**Correlation.**

$$\operatorname{Corr}(X,Y) \;\equiv\; \rho \;\equiv\; \frac{\operatorname{Cov}(X,Y)}{\sqrt{\operatorname{Var}(X)\operatorname{Var}(Y)}}, \qquad -1 \le \rho \le 1.$$

### 1.7 Conditional expectation and the tower property

$$\mathbb{E}[Y \mid X = x] \;\equiv\; \begin{cases} \sum_y y\, f(y \mid x) & \text{discrete,} \\ \int_{\mathbb{R}} y\, f(y \mid x)\, dy & \text{continuous.} \end{cases}$$

The result is a *function of* $x$, so $\mathbb{E}[Y \mid X]$ is itself a random variable.

**Property** (Tower / iterated expectation)**.**

$$\mathbb{E}\big[\,\mathbb{E}[Y \mid X]\,\big] \;=\; \mathbb{E}[Y].$$

This gives two ways to compute $\mathbb{E}[Y]$:

$$\mathbb{E}[Y] \;=\; \int_{\mathbb{R}} y\, f_Y(y)\, dy \;=\; \int_{\mathbb{R}} \mathbb{E}[Y \mid X = x]\, f_X(x)\, dx.$$

**Example 1.10.** With $f(x,y) = \frac{21}{4}x^2y$ on $x^2 \le y \le 1$, we computed $f(y \mid x) = 2y/(1-x^4)$. Then

$$\mathbb{E}[Y \mid X = x] \;=\; \int_{x^2}^{1} \frac{2y^2}{1-x^4}\, dy \;=\; \frac{2}{3} \cdot \frac{1-x^6}{1-x^4}.$$

At $x = 0.5$ this evaluates to about 0.70.

### 1.8 Random samples

$X_1, \dots, X_n$ is a *random sample* iff (i) the $X_i$ are independent and (ii) each has the same pdf/pmf $f(x)$, i.e., $X_i \overset{\text{iid}}{\sim} f(x)$.

For the sample mean $\bar{X} = \frac{1}{n}\sum X_i$ with $\mathbb{E}[X_i] = \mu$, $\operatorname{Var}(X_i) = \sigma^2$,

$$\mathbb{E}[\bar{X}] = \mu, \qquad \operatorname{Var}(\bar{X}) = \frac{\sigma^2}{n}.$$

The variance decreases as $n$ grows — this is what makes $\bar{X}$ a good estimator of $\mu$.

## 2 LOTUS, Moments, and Moment Generating Functions

### 2.1 LOTUS

**Law of the Unconscious Statistician.** If $Y = h(X)$ for a continuous $X$ with pdf $f(x)$,

$$\mathbb{E}[Y] \;=\; \int_{\mathbb{R}} h(x)\, f(x)\, dx.$$

You do not need the pdf of $Y$ to compute $\mathbb{E}[Y]$ — only the pdf of $X$ and the function $h$.

The $k$-th moment is $\mathbb{E}[X^k] = \int_{\mathbb{R}} x^k f(x)\, dx$, and the $k$-th central moment is $\mathbb{E}[(X-\mu)^k] = \int_{\mathbb{R}} (x-\mu)^k f(x)\, dx$. The variance is the second central moment:

$$\operatorname{Var}(X) \;=\; \mathbb{E}[(X-\mu)^2] \;=\; \mathbb{E}[X^2] - (\mathbb{E}[X])^2.$$

## 2.2 LOTUS via Taylor expansion (for messy $h$)

When $h$ is complicated, expand around $\mu = \mathbb{E}[X]$ with $\sigma^2 = \mathrm{Var}(X)$:

$$Y = h(X) \;\approx\; h(\mu) + (X-\mu)h'(\mu) + \tfrac{1}{2}(X-\mu)^2 h''(\mu) + R.$$

Taking expectations,

$$\boxed{\mathbb{E}[Y] \;\approx\; h(\mu) + \tfrac{1}{2}\, h''(\mu)\, \sigma^2.}$$

For variance, keep only the leading term:

$$\mathrm{Var}(Y) \;\approx\; \big[h'(\mu)\big]^2 \sigma^2.$$

## 2.3 Moment generating functions

**Definition 2.1.** The moment generating function (MGF) of $X$ is

$$M_X(t) \;=\; \mathbb{E}[e^{tX}].$$

$M_X$ is a function of $t$, not of $X$.

**Example 2.2** (Bernoulli MGF)**.** $X \sim \mathrm{Bern}(p)$ with $P(X=1)=p$, $P(X=0)=q=1-p$. Then

$$M_X(t) \;=\; e^{t\cdot 0}q + e^{t\cdot 1}p \;=\; q + pe^t.$$

**Why MGFs?** Under mild conditions,

$$\mathbb{E}[X^k] \;=\; \frac{d^k}{dt^k} M_X(t)\bigg|_{t=0}.$$

"The $k$-th moment is the $k$-th derivative of the MGF at $t=0$."

**Example 2.3** (Exponential)**.** If $X \sim \mathrm{Exp}(\lambda)$ then $M_X(t) = \lambda/(\lambda - t)$ for $t < \lambda$. Differentiating,

$$\mu = \mathbb{E}[X] = \frac{1}{\lambda}, \quad \mathbb{E}[X^2] = \frac{2}{\lambda^2}, \quad \mathrm{Var}(X) = \frac{1}{\lambda^2}.$$

## 2.4 MGF of a linear transformation

If $Y = aX + b$,

$$M_Y(t) \;=\; \mathbb{E}[e^{tY}] \;=\; \mathbb{E}[e^{atX+bt}] \;=\; e^{bt}\,\mathbb{E}[e^{(at)X}] \;=\; e^{bt}\, M_X(at).$$

**Example 2.4.** $X \sim \mathrm{Exp}(\lambda)$, $Y = 3X+2$. Then $M_Y(t) = e^{2t}\, M_X(3t) = e^{2t}\, \dfrac{\lambda}{\lambda - 3t}$.

MGFs uniquely identify distributions, so recognizing the form of $M_Y$ tells you what $Y$ is.

### 2.5 MGFs of sums

If $X_1, \ldots, X_n$ are *independent* and $Y = \sum X_i$,

$$M_Y(t) \;=\; \prod_{i=1}^{n} M_{X_i}(t).$$

If they are also identically distributed,

$$M_Y(t) \;=\; \bigl[M_{X_1}(t)\bigr]^n.$$

This is how we prove:

$$\sum \mathrm{Bern}(p) = \mathrm{Bin}(n,p), \quad \sum \mathrm{Geom}(p) = \mathrm{NegBin}(r,p), \quad \sum \mathrm{Exp}(\lambda) = \mathrm{Erlang}_n(\lambda), \quad \sum \mathrm{Pois}(\lambda_i) = \mathrm{Pois}\Bigl(\sum \lambda_i\Bigr)$$

## 3 Functions of Random Variables: Finding the Full Distribution

LOTUS gives $\mathbb{E}[Y]$ only — not the full distribution. To get the pdf/pmf of $Y = h(X)$:

### 3.1 Discrete $X$

If $X$ is discrete, so is $Y$, and

$$p_Y(y) \;=\; P(Y=y) \;=\; P\bigl(h(X)=y\bigr) \;=\; \sum_{x:\,h(x)=y} f_X(x).$$

In practice: build a table.

### 3.2 Continuous $X$ (CDF method)

Whether $Y$ is discrete or continuous, compute the CDF and then differentiate:

$$G(y) \;=\; P(Y \le y) \;=\; P\bigl(h(X) \le y\bigr) \;=\; \int_{\{x:\,h(x)\le y\}} f(x)\,dx,$$

then $g(y) = \frac{d}{dy} G(y)$.

## 4 The Normal Distribution

$X \sim \mathcal{N}(\mu, \sigma^2)$ iff

$$f(x) \;=\; \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right], \quad x \in \mathbb{R}.$$

The CDF $F(x)$ has no closed form. Key facts:

- $\mathbb{E}[X] = \mu$, $\mathrm{Var}(X) = \sigma^2$.
- $M_X(t) = \exp(\mu t + \frac{1}{2}\sigma^2 t^2)$.
- $P(\mu - \sigma < X < \mu + \sigma) = 0.6827$.
- $P(\mu - 2\sigma < X < \mu + 2\sigma) = 0.9545$.
- $P(\mu - 3\sigma < X < \mu + 3\sigma) = 0.9973$.

### 4.1 Additive property

If $X_1, \ldots, X_n$ are independent with $X_i \sim \mathcal{N}(\mu_i, \sigma_i^2)$ and constants $a_i, b_i$,

$$\boxed{Y = \sum_{i=1}^{n}(a_i X_i + b_i) \;\sim\; \mathcal{N}\Big(\sum a_i\mu_i + b_i,\ \sum a_i^2\sigma_i^2\Big).}$$

Variances always add; constants shift the mean only. Proof is by MGFs.

**Example 4.1.** $X \sim \mathcal{N}(3,4)$, $Y \sim \mathcal{N}(4,6)$ independent. Then

$$2X - 3Y \;\sim\; \mathcal{N}\big(2(3) - 3(4),\;\; 2^2(4) + (-3)^2(6)\big) \;=\; \mathcal{N}(-6, 70).$$

In particular, $X \sim \mathcal{N}(\mu, \sigma^2) \implies aX + b \sim \mathcal{N}(a\mu + b,\ a^2\sigma^2)$.

### 4.2 Standard normal

$$Z \;\equiv\; \frac{X - \mu}{\sigma} \;\sim\; \mathcal{N}(0,1).$$

The standard normal has pdf $\varphi(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}$ and CDF $\Phi(z)$. Useful identities:

$$\Phi(0) = \tfrac{1}{2}, \quad \Phi(-b) = 1 - \Phi(b), \quad P(-b \le Z \le b) = 2\Phi(b) - 1.$$

"Within $k$ standard deviations of the mean" does not depend on $\mu$ or $\sigma$:

$$P(\mu - k\sigma \le X \le \mu + k\sigma) \;=\; P(-k \le Z \le k) \;=\; 2\Phi(k) - 1.$$

### 4.3 Sample mean and the central limit theorem

If $X_i \overset{\text{iid}}{\sim} \mathcal{N}(\mu, \sigma^2)$, then by the additive property

$$\bar{X} \;\sim\; \mathcal{N}\left(\mu,\ \frac{\sigma^2}{n}\right).$$

The mean is constant in $n$; the variance shrinks. This is the Law of Large Numbers in action.

**Theorem** (Central Limit Theorem). If $X_1, \ldots, X_n$ are iid with mean $\mu$ and finite variance $\sigma^2$, then for large $n$,

$$\bar{X} \;\overset{d}{\approx}\; \mathcal{N}\left(\mu,\ \frac{\sigma^2}{n}\right), \quad \text{equivalently} \quad Z_n \;\equiv\; \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \;\xrightarrow{d}\; \mathcal{N}(0,1).$$

The original distribution of $X_i$ doesn't matter. Rule of thumb: $n \ge 30$ usually suffices; symmetric distributions need fewer samples.

# 5 Catalogue of Distributions

### 5.1 Discrete distributions

**Bernoulli,** Bern($p$). $P(X = 1) = p$, $P(X = 0) = q = 1 - p$. $\mathbb{E}[X] = p$, $\mathrm{Var}(X) = pq$, $M_X(t) = q + pe^t$.

**Binomial,** $\mathrm{Bin}(n,p)$**.** "$k$ successes in $n$ trials."

$$P(Y=k)=\binom{n}{k}p^k q^{n-k}, \quad k=0,1,\dots,n.$$

$\mathbb{E}[Y]=np$, $\mathrm{Var}(Y)=npq$, $M_Y(t)=(q+pe^t)^n$. A binomial is a sum of $n$ iid Bernoullis.

**Geometric,** $\mathrm{Geom}(p)$**.** *Number of trials until the first success.* Memoryless.

$$P(Z=k)=q^{k-1}p, \quad k=1,2,\dots \qquad P(Z>k)=q^k, \quad P(Z\le k)=1-q^k.$$

$\mathbb{E}[Z]=1/p$, $\mathrm{Var}(Z)=q/p^2$, $M_Z(t)=pe^t/(1-qe^t)$ for $t<\ln(1/q)$.

**Negative binomial,** $\mathrm{NegBin}(r,p)$**.** *Number of trials until the $r$-th success.*

$$P(W=k)=\binom{k-1}{r-1}p^r q^{k-r}, \quad k=r,r+1,\dots$$

$\mathbb{E}[W]=r/p$, $\mathrm{Var}(W)=rq/p^2$, $M_W(t)=\big(pe^t/(1-qe^t)\big)^r$.

**Hypergeometric.** $a$ of type 1 and $b$ of type 2; draw $n$ without replacement. "Like binomial but lower variance."

$$P(X=k)=\frac{\binom{a}{k}\binom{b}{n-k}}{\binom{a+b}{n}}, \quad \mathbb{E}[X]=n\cdot\frac{a}{a+b}, \quad \mathrm{Var}(X)=n\cdot\frac{a}{a+b}\left(1-\frac{a}{a+b}\right)\left(\frac{a+b-n}{a+b-1}\right).$$

**Poisson,** $\mathrm{Pois}(\lambda)$**.** $\lambda$ is the average number of occurrences per unit time/volume.

$$P(X=k)=\frac{\lambda^k e^{-\lambda}}{k!}, \quad k=0,1,2,\dots$$

$\mathbb{E}[X]=\lambda$, $\mathrm{Var}(X)=\lambda$, $M_X(t)=\exp(\lambda(e^t-1))$.

**Additive property.** Independent $X_i\sim\mathrm{Pois}(\lambda_i) \implies \sum X_i\sim\mathrm{Pois}(\sum\lambda_i)$. Always adjust $\lambda$ to match the time unit of the question.

**What is *not* Poisson:** customers arriving in groups; rate changing with time (e.g., dinner rush); arrivals that are not independent.

### 5.2 Continuous distributions

**Uniform,** $\mathrm{Unif}(a,b)$**.**

$$f(x)=\frac{1}{b-a},\ a\le x\le b, \quad F(x)=\frac{x-a}{b-a}, \quad \mathbb{E}[X]=\frac{a+b}{2}, \quad \mathrm{Var}(X)=\frac{(b-a)^2}{12},$$

$M_X(t)=(e^{tb}-e^{ta})/(t(b-a))$.

**Exponential,** $\mathrm{Exp}(\lambda)$**.** Memoryless, continuous analogue of geometric.

$$f(x)=\lambda e^{-\lambda x},\ x>0, \quad F(x)=1-e^{-\lambda x}, \quad \mathbb{E}[X]=1/\lambda, \quad \mathrm{Var}(X)=1/\lambda^2, \quad M_X(t)=\lambda/(\lambda-t).$$

Memoryless property: $P(X>s+t\mid X>s)=P(X>t)=e^{-\lambda t}$.

**Example 5.1.** A bulb has average life 10 months. Made it 20 months. Probability it makes 8 more? $\lambda=1/10$, and

$$P(X>28\mid X>20)\ =\ P(X>8)\ =\ e^{-8/10}\ =\ e^{-0.8}.$$

**Erlang,** $\text{Erlang}_k(\lambda)$**.** Sum of $k$ iid $\text{Exp}(\lambda)$ random variables.

$$f(s) = \frac{\lambda^k e^{-\lambda s} s^{k-1}}{(k-1)!}, \quad F(s) = 1 - \sum_{i=0}^{k-1} \frac{e^{-\lambda s}(\lambda s)^i}{i!}, \quad \mathbb{E}[S] = k/\lambda, \quad \text{Var}(S) = k/\lambda^2.$$

**Example 5.2.** $X, Y$ iid $\text{Exp}(2)$, find $P(X + Y \le 1)$. With $k = 2$, $s = 1$, $\lambda = 2$:

$$P(X + Y \le 1) \ = \ 1 - \sum_{i=0}^{1} \frac{e^{-2} \cdot 2^i}{i!} \ \approx \ 0.594.$$

**Gamma**$(\alpha, \lambda)$**.** Generalizes Erlang to non-integer shape $\alpha$.

$$f(x) = \frac{\lambda^\alpha e^{-\lambda x} x^{\alpha-1}}{\Gamma(\alpha)}, \quad \Gamma(\alpha) = \int_0^\infty t^{\alpha-1} e^{-t}\, dt, \quad \mathbb{E}[X] = \alpha/\lambda, \quad \text{Var}(X) = \alpha/\lambda^2.$$

$\Gamma(\alpha) = (\alpha - 1)!$ for positive integer $\alpha$; $\Gamma(1/2) = \sqrt{\pi}$.

**Triangular,** $\text{Tri}(a, b, c)$**.** Use to model RVs based on limited data (min, mode, max).

$$f(x) = \begin{cases} \dfrac{2(x-a)}{(b-a)(c-a)} & a \le x \le b, \\ \dfrac{2(c-x)}{(c-b)(c-a)} & b \le x \le c, \\ 0 & \text{otherwise.} \end{cases} \qquad \mathbb{E}[X] = \frac{a+b+c}{3}.$$

Variance is a mess.

**Beta**$(a, b)$**.** Use for RVs bounded on $(0, 1)$.

$$f(x) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} x^{a-1}(1-x)^{b-1},\ 0 < x < 1, \quad \mathbb{E}[X] = \frac{a}{a+b}, \quad \text{Var}(X) = \frac{ab}{(a+b)^2(a+b+1)}.$$

Very flexible: can reduce to Exp/Unif/Normal shapes. Can be rescaled to any interval $(a, b)$.

**Weibull, scale** $a$**, shape** $b$**.** Reliability models.

$$f(x) = ab(ax)^{b-1} e^{-(ax)^b},\ x > 0, \quad F(x) = 1 - \exp\left[-(ax)^b\right].$$

$\mathbb{E}[X] = (1/a)\Gamma(1 + 1/b)$. Variance is messy.

**Cauchy.**

$$f(x) = \frac{1}{\pi(1 + x^2)}, \quad F(x) = \frac{1}{2} + \frac{\arctan x}{\pi}, \quad x \in \mathbb{R}.$$

Mean and variance are undefined / infinite. **Weird:** if $X_i \overset{\text{iid}}{\sim}$ Cauchy, then $\bar{X} \sim$ Cauchy — the CLT does *not* apply.

### 5.3 Distribution relationships

- $\sum \mathrm{Bern}(p) = \mathrm{Bin}(n, p)$.
- $\sum \mathrm{Geom}(p) = \mathrm{NegBin}(r, p)$.
- Hypergeometric is binomial without replacement (lower variance).
- Poisson is the limit of $\mathrm{Bin}(n, p)$ as $n \to \infty$, $p \to 0$, $np \to \lambda$.
- Inter-arrival times of a Poisson process are $\mathrm{Exp}(\lambda)$.
- $\sum_{i=1}^{k} \mathrm{Exp}(\lambda) = \mathrm{Erlang}_k(\lambda)$. Erlang is a special case of Gamma.
- $\chi^2(2) \sim \mathrm{Exp}(1/2)$. $\sum_{i=1}^{n} Z_i^2 = \chi^2(n)$ where $Z_i \sim \mathcal{N}(0, 1)$.
- For $Z \sim \mathcal{N}(0, 1)$, $Y \sim \chi^2(k)$, $Z/\sqrt{Y/k} \sim t(k)$. $t(1) \sim$ Cauchy.
- $X \sim \chi^2(n)$, $Y \sim \chi^2(m) \implies (X/n)/(Y/m) \sim F(n, m)$.
- $U_1 + U_2 \sim \mathrm{Tri}(0, 1, 2)$ for $U_i \overset{\text{iid}}{\sim} \mathrm{Unif}(0, 1)$.
- Beta with $a = b = 1$ is $\mathrm{Unif}(0, 1)$.

## 6 Combinatorics and CDF Properties

### 6.1 Combinatorial facts

**Binomial theorem.** $(x + y)^n = \sum_{i=0}^{n} \binom{n}{i} x^i y^{n-i}$. Taking $x = y = 1$: $2^n = \sum_i \binom{n}{i}$.

**Hypergeometric (without replacement) vs binomial (with replacement).**

$$P(k \text{ type-A in } n \text{ draws, w/o repl}) = \frac{\binom{a}{k}\binom{b}{n-k}}{\binom{a+b}{n}},$$

$$P(k \text{ type-A in } n \text{ draws, with repl}) = \binom{n}{k}\left(\frac{a}{a+b}\right)^k \left(\frac{b}{a+b}\right)^{n-k}.$$

**Multinomial arrangements.**

$$\binom{n}{n_1, n_2, \ldots, n_k} = \frac{n!}{n_1!\, n_2! \cdots n_k!}.$$

**Example.** “MISSISSIPPI” has $\frac{11!}{1!\,4!\,4!\,2!} = 34{,}650$ arrangements.

**Disjoint vs. independent.** Not the same! Disjoint means $A$ and $B$ cannot co-occur (so knowing $A$ happened tells you $B$ didn’t, which is strong dependence). Independent means $P(A \cap B) = P(A)P(B)$.

### 6.2 CDF properties

For any CDF $F$:

1. Non-decreasing: $a < b \implies F(a) \leq F(b)$.
2. $\lim_{x \to \infty} F(x) = 1$, $\lim_{x \to -\infty} F(x) = 0$.
3. Right-continuous.
4. $P(X > x) = 1 - F(x)$.
5. $P(a < X \leq b) = F(b) - F(a)$.

**Linearity of $\mathbb{E}$ and $\mathrm{Var}$.**

$$\mathbb{E}[aX + b] = a\mathbb{E}[X] + b, \qquad \mathrm{Var}(aX + b) = a^2 \,\mathrm{Var}(X).$$

The constant "shifts" the mean but does not affect the variance.

**Proof: $\mathrm{Var}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2$.**

$$\begin{aligned} \mathrm{Var}(X) &= \mathbb{E}[(X - \mu)^2] \;=\; \mathbb{E}[X^2 - 2\mu X + \mu^2] \\ &= \mathbb{E}[X^2] - 2\mu\mathbb{E}[X] + \mu^2 \;=\; \mathbb{E}[X^2] - \mu^2. \end{aligned}$$

**Law of total probability.** $P(B) = P(A)P(B \mid A) + P(\bar{A})P(B \mid \bar{A})$.

### 6.3 Median and conditional probabilities

The median $m$ satisfies $F(m) = 0.5$. Conditional probabilities follow $P(A \mid B) = P(A \cap B)/P(B)$. Continuous example:

$$P\Big(0 < X < 1 \,\Big|\, \tfrac{1}{2} < X < \tfrac{3}{2}\Big) \;=\; \frac{P(\frac{1}{2} < X < 1)}{P(\frac{1}{2} < X < \frac{3}{2})} \;=\; \frac{\int_{1/2}^{1} f(x)\,dx}{\int_{1/2}^{3/2} f(x)\,dx}.$$

### 6.4 Failure rate function

For a continuous lifetime $X$ with pdf $f$ and CDF $F$, the (instantaneous) failure rate is

$$\zeta(t) \;=\; \frac{f(t)}{P(X > t)} \;=\; \frac{f(t)}{1 - F(t)}.$$

"Instantaneous rate of death after surviving to time $t$."

## 7 Hand Simulation, Euler's Method, and MSE

### 7.1 Euler's method

From $f'(x) = \lim_{h \to 0}(f(x + h) - f(x))/h$, for small $h$,

$$f(x + h) \;\approx\; h\,f'(x) + f(x).$$

**Example 7.1.** $f'(x) = (x + 1)f(x)$ with $f(0) = 1$, $h = 0.01$, $x \in [0, 0.2]$. Iterate:

$$f(x + h) \;\approx\; f(x) + h(x + 1)f(x).$$

### 7.2 Mean squared error

$$\mathrm{MSE}(T) \;\equiv\; \mathbb{E}\big[(T-\theta)^2\big] \;=\; \mathrm{Var}(T) + \big(\mathrm{Bias}(T)\big)^2, \quad \text{where } \mathrm{Bias}(T) = \mathbb{E}[T] - \theta.$$

We prefer estimators with low variance even if they are slightly biased — a low-variance, slightly biased estimator beats a high-variance unbiased one for most practical purposes.

## 8 Poisson Processes

A Poisson process $\{N(t), t \geq 0\}$ is a counting process with arrival rate $\lambda$ (average occurrences per unit time/volume). Three assumptions:

1. Arrivals happen one at a time, with rate $\lambda$ per unit time.
2. The arrival pattern is stationary.
3. Counts in disjoint time intervals are independent.

Inter-arrival times are iid $\mathrm{Exp}(\lambda)$.

**Sums of distributions.** If $U_1, U_2 \sim \mathrm{Unif}(0,1)$ then $U_1 + U_2 \sim \mathrm{Tri}(0,1,2)$. And $-\frac{1}{\lambda}\ln(U) \sim \mathrm{Exp}(\lambda)$, where $U \sim \mathrm{Unif}(0,1)$ — this is the inverse-transform method.

**Sample mean and CLT (recap).** $\bar{X} \sim \mathcal{N}(\mu, \sigma^2/n)$ when $X_i \overset{\text{iid}}{\sim} \mathcal{N}(\mu, \sigma^2)$, and the CLT extends this approximately to any iid distribution with finite variance.

## 9 Simulation with Arena

### 9.1 Process-interaction model

Arena uses the process-interaction worldview. *Entities* (with events and activities) flow through *modules* (blocks). Many entities interact with each other in the system.

**Basic Process Template.** Most-used template. Building blocks:

- **Create** (entity arrivals)
- **Process** (resource + queue — automatically sets up a queue named "{ProcessName}.Queue")
- **Dispose**
- **Batch**, **Decide**, **Separate**, **Assign**, **Record**.

**Spreadsheets** accompany the modules: entity attributes, resource capacities, schedules, etc.

### 9.2 The Process module

Grabs servers, uses them, lets them go. Sets up a queue automatically. Action types:

- **Seize** (self-service).
- **Seize-Delay-Release** (the common case — formal name).

- **Seize-Delay** (e.g., hospital beds: release happens elsewhere after cleanup). **Will deadlock if no Release later!**
- **Delay-Release**: releases a previously-seized server.

**Resource spreadsheet.** *Capacity is the number of resources available, not the number of customers requesting them.* Resources are auto-created by the Process module.

**Schedule spreadsheet.** Change capacity over time (e.g., servers taking breaks), change customer arrival patterns.

## 9.3 Decide and Assign modules

**Decide.**

- Chance (2-way or $n$-way).
- Condition (2-way or $n$-way).

Example: send to P1 if $NQ(P1.Queue) < NQ(P2.Queue)$, else P2.

**Assign.** Sets attributes on the entity (numerical only, but you can use "pictures"). Attributes are specific to the entity; *variables are global.* Results go to spreadsheets.

## 9.4 Internal variables

Global, affect the entire simulation:

- `TNOW` — current simulation time.
- `NR(Resource)` — number of resource servers currently working (not the total!).
- `NQ(Process.Queue)` — number of customers in queue.
- `{Module}.NumberOut` — count of entities that have exited.

## 9.5 Batch, Separate, Record

**Batch.** Groups customers into "super customers" (useful for non-homogeneous Poisson modeling). *Permanent* batching makes the super-customer lose individual attributes.

**Separate.** Splits a batch *or* duplicates customers.

**Record.** Just records statistics. Simple.

## 9.6 Run setup

Standard configuration: replications, warm-up time, accumulate-stats / re-initialize, speed, terminating condition, "Category Overview" for summary output. Turn off graphics with Batch Mode (fast!).

## 9.7 Advanced Arena

- Attributes are “memorized” by incoming customers — use Assign.
- Logical expressions go in parentheses: `IF TRUE` → `yield 1; ELSE` → `yield 0`.
- “Fake customers” — don’t care about their service times. Use them to stress-test or generate breakdowns.
- Advanced Process Template ships with separate Seize / Delay / Release modules, allowing Seize / Assign / Delay / Release and asymmetric Seize/Release patterns and complex Seize rules.

## 9.8 Resource failures

Can use “fake customers” to cause failures, or use the Resource spreadsheet (Basic Template), or the Failure spreadsheet (Advanced Template). Failure parameters:

- Count (#).
- Time (duration of).
- Downtime (distribution).

**Failure rules.**

- **Ignore**: serve customer, then reduce remaining repair time. (Service 10 min, repair 1 hr; if failure occurs at 60 min, repair = 50 min.)
- **Wait**: serve customer, then delay repair. (Finishes service at 70 min, then full repair.)
- **Preempt**: stop everything, repair, then serve customer.

## 9.9 Resource sets

Make a set of resources where each element has different abilities. You can Seize with rules: cyclical, random, “most resources in set,” and many more. *You have to release the same server* — remember the index of the agent that served.

## 9.10 Queue configuration

One-line vs. two-line: one combined line is almost always better.

## 9.11 Advanced Manufacturing Cell Demo

Uses the Advanced Transfer template. Makes 3 kinds of parts; each follows a different *sequence* through service stations (e.g., S1 → S2 → S1 → S3 → S4 → S2).

- Need a sequences spreadsheet.
- Need Route, Station, Enter, Leave modules from the Advanced Transfer template.
- Need Advanced Sets (of sequences). You can make sets of anything.
- Use the Route block to define sequences (defined in the sequences spreadsheet): “Go from X to Y in $Z$ time.”

- Use Station to define destinations.

**Transporters** carry customers between stations — resources that are movable (forklifts, cabs, etc.).

**Conveyors** move things along belts, elevators, etc.

### 9.12 SMARTS files

are Rockwell-provided tutorials (not just demos). Re-entrant queues in lectures oscillate and get larger over time.

## 10 Module 6: Random Number Generation

### 10.1 Goal and desirable properties

Goal: $\text{Unif}(0,1)$ inputs $\to$ Exp/Normal/etc. outputs. Uniform random variables are the key to random variate generation.

Desirable PRNG properties:

- Output appears to be iid.
- Fast.
- Reproducible sequence (deterministic — you can loop through, save state).

### 10.2 Bad generators

**Coin/die tosses.** OK but storage is a problem (we are generating billions).

**Book of random digits.** "Random" in print but no longer random once printed. Too small.

**Mid-square method (von Neumann).**

$$X_0 = 6632 \to 6632^2 = 43,\underline{9534}24 \to X_1 = 9534; \quad 9534^2 = 90,\underline{8975}56 \to X_2 = 8975; \ \ldots$$

$R_i = X_i/10{,}000$. **Problems:** positive serial correlation; can hit "0003" — "degenerates"!

**Fibonacci / additive congruential.**

$$X_i = (X_{i-1} + X_{i-2}) \bmod m.$$

Problems: small numbers follow small numbers (positive serial correlation); can't get "stuff in between" (e.g., never $X_2 < X_3 < X_1$ with probability 1/3).

### 10.3 Linear Congruential Generator (LCG)

General form:

$$X_i \ = \ (aX_{i-1} + c) \bmod m, \qquad R_i = X_i/m.$$

Pick $a, c, m$ to get good statistical properties and good cycle length. If the cycle length equals $m$, we have a *full-period generator*.

**Hyperplane issue.** LCG-generated points fall on lines/planes: plot $(X_{i-1}, X_i)$ and you see structure. The infamous example is RANDU,

$$X_i = 65{,}539\, X_{i-1} \bmod 2^{31},$$

which produces only *15 hyperplanes* in 3D.

**Desert Island generator.**

$$X_i = 16807\, X_{i-1} \bmod (2^{31} - 1), \quad R_i = X_i/(2^{31} - 1).$$

Multiplicative ($c = 0$). Full cycle if you do not start at zero.

**Algorithm (avoid overflow).**

$$\begin{aligned} X_i &= 16807[X_{i-1} - 127{,}773 \cdot \lfloor X_{i-1}/127{,}773 \rfloor] - 2836\, \lfloor X_{i-1}/127{,}773 \rfloor, \\ &\quad \text{if } X_i < 0: \ X_i \leftarrow X_i + (2^{31} - 1), \\ &\quad R_i \leftarrow X_i \cdot (4.656612875 \times 10^{-10}). \end{aligned}$$

## 10.4 Where LCGs go wrong

- Not full-period: e.g., $(4X_{i-1} + 2) \bmod 8$ generates only even numbers.
- Full-period but not random: $X_i = (X_{i-1} + 1) \bmod 8$ with $X_0 = 0$ gives $0, 1, 2, 3, \ldots$ (LOL).
- $m$ too small $\implies$ quick cycling (below $\sim$ 2 billion).
- $m$ large but bad parameters give problems like 3D hyperplanes in RANDU.

## 10.5 Tausworthe generator

Provides a sequence of binary digits $B_i$:

$$B_i \;=\; \left( \sum_{j=1}^{q} c_j B_{i-j} \right) \bmod 2, \quad c_j \in \{0, 1\}.$$

Period is always $2^q - 1$. Simplified recursion: $B_i = B_{i-r} \oplus B_{i-q}$ (XOR), $0 < r < q$.

**Example.** $r = 3$, $q = 5$, initialize $B_1, \ldots, B_5 = 1$. Then $B_i = B_{i-3} \oplus B_{i-5}$ for $i > 5$:

$$B_6 = B_3 \oplus B_1 = 0, \quad B_7 = B_4 \oplus B_2 = 0, \ \ldots$$

$B_i = 0$ if $B_{i-r} = B_{i-q}$; $B_i = 1$ otherwise. Get uniforms by reading $\ell$ bits in base 2:

$$\text{e.g., } 1111_2/2^4 \;=\; 15/16.$$

## 10.6 Generalizations of LCGs

Simple multi-term:

$$X_i = \left( \sum_{j=1}^{q} a_j X_{i-j} \right) \bmod m.$$

Can have period up to $m^q - 1$ (not guaranteed). Fibonacci is a special case.

**Combination + shuffle.** Make $Z_i = (X_i + Y_i) \bmod m$, then shuffle: set $Z_i = X_i$ or $Y_i$. Properties hard to prove.

**L'Ecuyer (1999).** Nasty but gives cycle length of approximately $2^{191}$.

**Mersenne Twister.** Cycle length $2^{19937} - 1$. Typically we don't need PRNGs larger than $2^{100}$.

**$k$-tuples.** Means $k$ sequential/consecutive random numbers $(R_i, R_{i+1}, \ldots, R_{i+k-1})$.

### 10.7 Hyperplanes: design goal

Fill space with as many hyperplanes as possible to look like "real" noise. Equivalently:

- Maximize the number of planes.
- Minimize the distance between adjacent planes.
- Maximize the minimum distance between adjacent $k$-tuples.

**Serial correlation (Greenberger, 1961).**

$$\mathrm{Corr}(R_1, R_2) \;\le\; \frac{1}{a}\left[1 - \frac{6c}{m} + 6\left(\frac{c}{m}\right)^2\right] + \frac{a+6}{m}.$$

$m$'s upper bound has to be very small ($\approx 2$ billion).

## 11 Statistical Tests for PRNGs

Two kinds:

1. Goodness-of-fit: are the $R_i$ really $\sim \mathrm{Unif}(0,1)$?
2. Independence tests: are the $R_i$ independent?

### 11.1 $\chi^2$ goodness-of-fit test

Null hypothesis: $R_1, R_2, \cdots \sim \mathrm{Unif}(0,1)$.

- Significance level $\alpha = P(\text{reject } H_0 \mid H_0 \text{ true})$ — Type I error.
- $\beta = P(\text{accept } H_0 \mid H_0 \text{ false})$ — Type II error (more important!).

Slot $R_n$'s into $k$ cells. We want the expected number of $R_n$'s in each cell to match what the actual distribution predicts.

$$O_i \sim \mathrm{Bin}(n, 1/k), \qquad E_i = \mathbb{E}[O_i] = n/k.$$

Test statistic:

$$\chi_0^2 \;=\; \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i}.$$

Big value $\implies$ bad fit. Reject $H_0$ iff $\chi_0^2 > \chi^2_{\alpha,k-1}$.

**Approximation for huge $n$, $k$.**

$$\chi^2_{\alpha,k-1} \approx (k-1)\left[1 - \frac{2}{9(k-1)} + Z_\alpha\sqrt{\frac{2}{9(k-1)}}\right]^3.$$

**Other GoF tests.** Kolmogorov–Smirnov, Anderson–Darling, Shapiro–Wilk (for normal).

**Example 11.1.** $n = 1000$, $k = 5$ cells. Each $E_i = 200$; observed $O_i = 178, 292, 201, 172, 162$ (approx). Compute $\chi^2_0$, compare against $\chi^2_{0.05,4}$, decide.

### 11.2 Independence tests

**Runs.** A run is a series of consecutive identical observations. "HHHHTTTTT" has positive serial correlation; "HTHTHTHT" has negative.

**Up and down (FUN!).**

$$0.41,\ 0.68,\ 0.89,\ 0.84,\ 0.74,\ 0.91 \ \to\ +, +, -, -, + \ \Rightarrow\ A = 3 \text{ runs (of identical sign).}$$

For $n \geq 20$ iid samples, $A$ is approximately normal:

$$A \overset{d}{\approx} \mathcal{N}\left(\frac{2n-1}{3},\ \frac{16n-29}{90}\right).$$

Test statistic and rejection:

$$Z_0 = \frac{A - \mathbb{E}[A]}{\sqrt{\text{Var}(A)}}; \qquad \text{reject if } |Z_0| > z_{\alpha/2}.$$

**Above and below mean.** Mean is 0.5. Add + if $R_i \geq 0.5$, $-$ if $R_i < 0.5$. Let $B$ = number of runs. With $n = n_1 + n_2$ where $n_1 = \#$ above and $n_2 = \#$ below 0.5,

$$B \overset{d}{\approx} \mathcal{N}\left(\frac{2n_1n_2}{n} + \frac{1}{2},\ \frac{2n_1n_2(2n_1n_2 - n)}{n^2(n-1)}\right).$$

Same test stat $Z_0 = (B - \mathbb{E}[B])/\sqrt{\text{Var}(B)}$; reject if $|Z_0| > z_{\alpha/2}$.

## 12 Module 6–7: Random Variate Generation

Use $\text{Unif}(0,1)$ inputs to generate "realizations" from any other distribution. Key tools:

1. Inverse transform.
2. Convolution (sums).
3. Acceptance-rejection.
4. Composition.
5. Special algorithms (Box-Muller, polar, etc.).

### 12.1 Inverse transform theorem

**Key fact:** If $F$ is the CDF of $X$, then $F(X) \sim \text{Unif}(0,1)$. Setting $F(X) = U$ and solving for $X$ gives a sample.

**Exponential.** $F(x) = 1 - e^{-\lambda x} = U \implies X = -\frac{1}{\lambda}\ln(U)$ (equivalently $-\frac{1}{\lambda}\ln(1-U)$).

**Weibull.** $F(x) = 1 - e^{-(\lambda x)^\beta} = U \implies X = \frac{1}{\lambda}[-\ln(1-U)]^{1/\beta}$ (or with $U$).

**Triangular(0,1,2).**

$$f(x) = \begin{cases} x & 0 \le x \le 1 \\ 2-x & 1 \le x \le 2 \end{cases}, \qquad F(x) = \begin{cases} x^2/2 & 0 \le x \le 1 \\ 1-(x-2)^2/2 & 1 \le x \le 2 \end{cases}.$$

Split by $U$:

$$X = \sqrt{2U} \ \text{ if } U < \tfrac{1}{2}, \qquad X = 2 - \sqrt{2(1-U)} \ \text{ if } U \ge \tfrac{1}{2}.$$

**Care with signs!** $X$ has to be between 1 and 2 when $U \ge 1/2$. You cannot generally replace $U$ with $(1-U)$ when the distribution is asymmetric.

**Normal.** $\Phi^{-1}(\cdot)$ has no closed form. Do a table lookup: if $U = 0.975$, $Z = \Phi^{-1}(0.975) = 1.96$. To generate $\mathcal{N}(\mu, \sigma^2)$, use $X = \mu + \sigma Z$. Example: $\mathcal{N}(3, 16)$ from $U = 0.59$: $X = 3 + 4\Phi^{-1}(0.59) \approx 3.91$.

**Discrete distributions (Bernoulli, etc.).** Build a table of $P(X = x)$ and the CDF, then map ranges of $U$ to outcomes.

**Geometric.** $X = \lceil \ln(U)/\ln(1-p) \rceil = \lceil \ln(1-U)/\ln(1-p) \rceil$. Equivalently, count Bernoulli($p$) trials until the first success.

**Poisson.** Same idea as the "other" (table) method: list $P(X = x)$, accumulate to get the CDF, slot $U$ into the appropriate bin.

### 12.2 Empirical distributions

When you don't know the underlying distribution, use the empirical CDF

$$\hat{F}_n(x) \;=\; \frac{\#\{X_i \le x\}}{n},$$

a step function. Glivenko–Cantelli: $\hat{F}_n(x) \to F(x)$ as $n \to \infty$, so it works in the continuous context too.

**Linear interpolation for continuous use.** Sort to get order statistics $X_{(1)} \le X_{(2)} \le \cdots \le X_{(n)}$. Define

$$F_{\text{interp}}(x) = \begin{cases} 0 & x < X_{(1)}, \\ \dfrac{i-1}{n-1} + \dfrac{x - X_{(i)}}{(n-1)\left[X_{(i+1)} - X_{(i)}\right]} & X_{(i)} \le x \le X_{(i+1)}, \\ 1 & x \ge X_{(n)}. \end{cases}$$

To sample: set $F(X) = U$, let $P = (n-1)U$, $I = \lceil P \rceil$, then

$$X \;=\; X_{(I)} + (P - I + 1)\big[X_{(I+1)} - X_{(I)}\big].$$

**Example 12.1.** $X_{(1)} = 1,\ X_{(2)} = 4,\ X_{(3)} = 6$. With $U = 0.73$:

$$P = 2 \cdot 0.73 = 1.46,\ I = \lceil 1.46 \rceil = 2,\ X = X_{(2)} + (1.46 - 2 + 1)[X_{(3)} - X_{(2)}] = 4 + 0.46(2) = 4.92.$$

## 12.3 Convolution: sums of distributions

If $Y = \sum X_i$ where the $X_i$'s come from a known distribution, the distribution of $Y$ is often well-known. Convolution = adding distributions up:

- $\sum_{i=1}^{n} \text{Bern}(p) = \text{Bin}(n, p)$.
- $U_1 + U_2 \sim \text{Tri}(0, 1, 2)$.
- $\sum_{i=1}^{n} \text{Exp}(\lambda) = \text{Erlang}_n(\lambda)$, so $Y = -\frac{1}{\lambda} \ln\big(\prod_{i=1}^{n} U_i\big)$.
- $\sum \text{Geom} = \text{NegBin}$.
- $\sum Z_i^2 = \chi^2(n)$.
- Crude Desert Island: $Y = \sum_{i=1}^{12} U_i - 6 \approx \mathcal{N}(0,1)$ (by CLT, since $\mathbb{E}\sum U_i = 6$, $\text{Var}\sum U_i = 1$).
- **Cauchy violates the CLT**: $\bar{X} \sim$ Cauchy when $X_i \sim$ Cauchy.

## 12.4 Acceptance-rejection

Works for messy pdfs where the inverse transform doesn't yield a closed form. *Always works, somehow.*

**Setup.** Want to sample from $f(x)$. Find a *majorizer* $t(x) \geq f(x)$ for all $x$. Let

$$c = \int_{\mathbb{R}} t(x)\, dx, \quad h(x) = t(x)/c, \quad g(x) = f(x)/t(x).$$

**Algorithm.** Loop: generate $X$ from $h(x)$ and $U \sim \text{Unif}(0,1)$. If $U \leq g(X)$, accept; else reject and try again.

**Example 12.2.** $f(x) = 60x^3(1-x)^2$ on $[0,1]$. Calculus shows the max of $f$ is 2.0736 at $x = 0.6$. So use the dumb majorizer $t(x) = 2.0736$ — a flat line. Then $c = \int_0^1 t(x)\, dx = 2.0736$, $h(x) = 1$ (uniform), $g(x) = 60x^3(1-x)^2/2.0736$. To sample: $X \sim \text{Unif}(0,1)$, $U \sim \text{Unif}(0,1)$, accept if $U \leq g(X)$.

**Connection to Poisson.** Recall: $X = n$ if exactly $n$ arrivals occur in a unit time of a $\text{Pois}(\lambda)$ process, with inter-arrival times $A_i \sim \text{Exp}(\lambda)$. So

$$\sum_{i=1}^{n} A_i \leq 1 < \sum_{i=1}^{n+1} A_i \iff X = n.$$

Since $A_i = -\frac{1}{\lambda}\ln(U_i)$, this becomes

$$\boxed{\prod_{i=1}^{n} U_i \;\geq\; e^{-\lambda} \;>\; \prod_{i=1}^{n+1} U_i.}$$

TL;DR: multiply $U_i$ until the product is smaller than $e^{-\lambda}$.

**Example 12.3.** $\lambda = 2$, $e^{-2} \approx 0.1353$. Generate $U_1, U_2, \dots$ until $\prod U_i < 0.1353$.

Expected number of $U_i$'s needed: $\mathbb{E}[X+1] = \lambda + 1$. **For** $\lambda \geq 20$**,** use the normal approximation: $(X-\lambda)/\sqrt{\lambda} \approx \mathcal{N}(0,1)$, so $X = \max(0, \lfloor \lambda + \sqrt{\lambda}\, Z + 0.5 \rfloor)$.

## 12.5 Generating Uniforms on $(a, b)$

E.g., $\mathrm{Unif}(2/3, 1)$: roll until $U_i \geq 2/3$. Wasteful but easy.

## 12.6 Composition method

Useful for mixtures of RVs that are not the same distribution. If

$$F(x) \;=\; \sum_{j=1}^{\infty} p_j F_j(x), \qquad \sum p_j = 1,$$

then "with probability $p_j$, sample from $F_j$."

**Laplace example.**

$$X \;\leftarrow\; \begin{cases} \ln(U) & \text{w.p. } 1/2, \\ -\ln(U) & \text{w.p. } 1/2. \end{cases}$$

## 12.7 Box-Muller (Normal)

$U_1, U_2 \sim \mathrm{Unif}(0,1)$ independent. Then

$$Z_1 = \sqrt{-2\ln(U_1)}\,\cos(2\pi U_2), \qquad Z_2 = \sqrt{-2\ln(U_1)}\,\sin(2\pi U_2),$$

are iid $\mathcal{N}(0,1)$. **Angles must be in radians!**

## 12.8 Polar method (alternative to Box-Muller)

$U_1, U_2 \sim \mathrm{Unif}(0,1)$. Center: $V_i = 2U_i - 1$. Let $W = V_1^2 + V_2^2$. Loop: if $W > 1$, regenerate. Else,

$$Y = \sqrt{-2\ln(W)/W}, \qquad Z_i = V_i Y \sim \mathcal{N}(0,1).$$

## 12.9 Order statistics

If $Y = \min\{X_1, \dots, X_n\}$ where each $X_i$ has CDF $F$ and the $X_i$ are iid, then $Y$ has CDF

$$G(y) \;=\; 1 - P(Y > y) \;=\; 1 - P(\min(X_i) > y) \;=\; 1 - \big[P(X_1 > y)\big]^n \;=\; 1 - \big[1 - F(y)\big]^n.$$

Solve $G(Y) = U$ for $Y$:

$$Y \;=\; F^{-1}\Big[1 - (1-U)^{1/n}\Big].$$

**Example 12.4.** $X_1, \dots, X_n \overset{\text{iid}}{\sim} \mathrm{Exp}(\lambda)$, want the minimum from just one $U$. $G(y) = 1 - \big[1 - (1 - e^{-\lambda y})\big]^n = 1 - e^{-\lambda y n}$, which is the CDF of $\mathrm{Exp}(n\lambda)$. So $Y = -\frac{1}{n\lambda}\ln(U)$.

**Other order-statistic facts.**

$$\sum_{i=1}^{n} Z_i^2 \sim \chi^2(n), \qquad \frac{Z}{\sqrt{Y/n}} \sim t(n) \text{ when } Z \sim \mathcal{N}(0,1),\ Y \sim \chi^2(n).$$

$t(1) \sim$ Cauchy. With $X \sim \chi^2(n)$, $Y \sim \chi^2(m)$, $(X/n)/(Y/m) \sim F(n,m)$. $Z_2/Z_1 \sim$ Cauchy $\sim t(1)$. $Z_1^2 + Z_2^2 \sim \chi^2(2) \sim \text{Exp}(1/2)$.

## 12.10 Markov chains

You have a state-transition matrix; you walk through states based on the matrix. Easy to simulate with uniforms.

**Example 12.5.** Rain/Sun chain: $P(\text{Rain} \mid \text{Rain}) = 0.7$, $P(\text{Sun} \mid \text{Rain}) = 0.3$, $P(\text{Rain} \mid \text{Sun}) = 0.4$, $P(\text{Sun} \mid \text{Sun}) = 0.6$. If Monday is rain, simulate Tuesday: draw $U$, compare to 0.7, $\rightarrow$ rain or sun, etc.

## 12.11 Poisson processes (simulation)

**(1) Stream of times at rate $\lambda$.** $T_0 = 0$, $T_i = T_{i-1} + A_i$ where $A_i \sim \text{Exp}(\lambda)$, so $T_i = T_{i-1} - \frac{1}{\lambda}\ln(U)$.

**(2) Guaranteed $n$ arrivals in $[a,b]$.** There's a nice theorem: take $n$ uniforms $U_1, \ldots, U_n \sim \text{Unif}(a,b)$, sort them, done. Adjust: $T_i = a + (b-a)U_{(i)}$.

## 12.12 Non-homogeneous Poisson process

Arrival rate $\lambda(t)$ is a function. Generally,

$$N(s+t) - N(s) \ \sim\ \text{Pois}\left(\int_s^{s+t} \lambda(u)\, du\right).$$

**Example 12.6.** $\lambda(t) = t^2$. $N(2) - N(1) \sim \text{Pois}(\int_1^2 t^2\, dt) = \text{Pois}(7/3)$. Probability of exactly 4 arrivals:

$$P(N(2) - N(1) = 4) \ =\ \frac{e^{-7/3}(7/3)^4}{4!}.$$

**Thinning algorithm.** Find $\lambda^* = \max_t \lambda(t)$. Generate potential arrivals at rate $\lambda^*$ using $\text{Exp}(\lambda^*)$. For each candidate $t$, accept it with probability $\lambda(t)/\lambda^*$ (draw $V \sim \text{Unif}(0,1)$, accept if $V \leq \lambda(t)/\lambda^*$). **Note: $t$ is the running sum of all inter-arrivals so far, not the inter-arrival itself.**

## 12.13 Brownian motion

"Probably the most important stochastic process out there" — Brown / Einstein / Wiener. Actually used in queueing theory, finance, etc.

1. $W(0) = 0$.
2. $W(t) \sim \mathcal{N}(0,t)$.
3. $\{W(t), t \geq 0\}$ has stationary, independent increments.

**Stationary:** $W(t+h) - W(t)$ depends only on $h$. **Independent:** for $a < b \leq c < d$, $W(d) - W(c) \perp W(b) - W(a)$.

**Construction (Donsker's CLT).** $X_n \sim \mathcal{N}(0,1)$ iid; then

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{\lfloor nt \rfloor} X_i \xrightarrow{d} W(t).$$

Setting $t = 1$ recovers the ordinary CLT, $W(1) \sim \mathcal{N}(0,1)$.

**Generating numerically.** Successive: $W(0) \to W(1) \to W(2) \to \cdots$. The recursion is

$$W\left(\tfrac{i}{n}\right) = W\left(\tfrac{i-1}{n}\right) + \tfrac{Y_i}{\sqrt{n}},$$

where $Y_i$ is $\pm 1$ with probability $1/2$ (or $\mathcal{N}(0,1)$). $n$ needs to be large.

**Properties.**

1. Too squiggly — no derivatives exist!
2. $\mathrm{Cov}(W(s), W(t)) = \min(s,t)$.
3. $\int_0^1 W(t)\,dt \sim \mathcal{N}(0, 1/3)$.

**Brownian bridge $B(t)$.** Conditioned on $W(0) = W(1) = 0$.

(5) $\mathrm{Cov}(B(s), B(t)) = \min(s,t) - st$.

(6) $\int_0^1 B(t)\,dt \sim \mathcal{N}(0, 1/12)$.

**Geometric Brownian motion.** Stock prices,

$$S(t) = S(0)\,\exp\Big[\Big(\mu - \tfrac{\sigma^2}{2}\Big)t + \sigma W(t)\Big], \quad t \geq 0.$$

$S(0)$ = initial price, $\mu$ = drift, $\sigma$ = volatility. The $-\sigma^2/2$ is a volatility penalty: we only want stocks whose price drifts up, so we want $\mu - \sigma^2/2 > 0$.

**Option pricing.** European call: $e^{-rT}\,\mathbb{E}[(S(T) - K)^+]$, where $K$ = strike price, $T$ = expiry date, $r$ = risk-free rate. Solve this to win a Nobel prize.

### 12.14 Bivariate normal $k = 2$

Generate $\mathbf{X} = \boldsymbol{\mu} + C\mathbf{Z}$ where $\mathbf{Z} \sim \mathcal{N}(0, I)$ and $C$ is the (lower-triangular) Cholesky factor of the covariance matrix:

$$C = \begin{pmatrix} \sqrt{\sigma_{11}} & 0 \\ \sigma_{12}/\sqrt{\sigma_{11}} & \sqrt{\sigma_{22} - \sigma_{12}^2/\sigma_{11}} \end{pmatrix}.$$

So

$$X_1 = \mu_1 + \sqrt{\sigma_{11}}\,Z_1, \qquad X_2 = \mu_2 + \frac{\sigma_{12}}{\sqrt{\sigma_{11}}} Z_1 + \sqrt{\sigma_{22} - \frac{\sigma_{12}^2}{\sigma_{11}}}\,Z_2.$$

## 13 Module 8: Input Analysis

Input data can "GIGO" a simulation, so we need a good set of RVs for things like arrival times.

### 13.1 Workflow

1. Collect data.
2. Pick a distribution (hypothesis).
3. Perform goodness-of-fit tests; if (2) fits (1), OK; else iterate.

**Histograms first.** With enough observations the histogram converges to the true density (Glivenko–Cantelli). More observations = better, but too many can confound the choice.

**Distribution guidance.**

- Triangular: when you know min, max, and "most likely."
- Uniform: when you don't know much beyond min and max.
- Normal: heights, weights, IQs.
- Beta: bounded data.
- Gamma, Weibull, Gumbel, Lognormal: reliability data.

You can always bail and use the empirical / sample distribution.

### 13.2 Unbiased point estimation

$\bar{X}$ and $S^2$ are statistics; $\mu$ and $\sigma^2$ are parameters. A statistic $T[X]$ is unbiased for $\theta$ iff $\mathbb{E}[T[X]] = \theta$.

- For Exp($\lambda$): $\bar{X}$ is unbiased for $1/\lambda$, but $\lambda \neq 1/\bar{X}$ in expectation.
- For normal: $\mathbb{E}[S^2] = \sigma^2$ (so $S^2$ is unbiased for $\sigma^2$), but $\mathbb{E}[S] \neq \sigma$ (so $S$ is *biased* for $\sigma$).

**Good properties for $T$.**

1. Unbiased.
2. Low bias (high bias + low variance = noisy).
3. Low variance (low bias + high variance = overconfident).

$\mathrm{Bias}(T) = \mathbb{E}[T] - \theta$.

$$\mathrm{MSE}(T) = \mathbb{E}[(T-\theta)^2] = \mathrm{Var}(T) + \mathrm{Bias}(T)^2.$$

Lower MSE is better even if there is some bias.

**Relative efficiency.** $RE(T_1, T_2) = \mathrm{MSE}(T_1)/\,\mathrm{MSE}(T_2)$. If $RE < 1$, pick $T_1$.

### 13.3 Maximum likelihood estimation

Most popular point-estimator approach. Has some bias but is OK on average. Likelihood

$$L(\theta) \;\equiv\; \prod_{i=1}^{n} f(x_i \mid \theta).$$

The MLE $\hat{\theta}$ maximizes $L(\theta)$. Use the log-likelihood: $\ln L$ is one-to-one and easier to differentiate.

**Example 13.1** (Exponential MLE)**.** $X_i \overset{\text{iid}}{\sim} \text{Exp}(\lambda)$. $L(\lambda) = \prod \lambda e^{-\lambda x_i} = \lambda^n \exp\big(-\lambda \sum x_i\big)$.

$$\ln L = n \ln \lambda - \lambda \sum x_i \;\Rightarrow\; \frac{d}{d\lambda} \ln L = \frac{n}{\lambda} - \sum x_i = 0 \;\Rightarrow\; \hat{\lambda} = \frac{1}{\bar{X}}.$$

**Example 13.2** (Bernoulli MLE)**.** $\hat{p} = \bar{X}$.

**Example 13.3** (Normal MLE)**.** $X_i \overset{\text{iid}}{\sim} \mathcal{N}(\mu, \sigma^2)$. Take partial derivatives of $\ln L(\mu, \sigma^2)$:

$$\hat{\mu} = \bar{X}, \qquad \hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^{n} (X_i - \bar{X})^2.$$

Note $\hat{\sigma}^2 \neq S^2$ (which uses $n-1$ in the denominator). $S^2$ is unbiased; $\hat{\sigma}^2$ is biased but has lower variance.

**Example 13.4** (Uniform on $(0, \theta)$)**.** $L(\theta) = (1/\theta)^n$ if $0 \le X_i \le \theta$ for all $i$, else 0. $L$ is maxed at the *smallest* allowable $\theta$:

$$\hat{\theta} = \max_i X_i.$$

### 13.4 Invariance property of MLE

If $\hat{\theta}$ is the MLE of $\theta$ and $h$ is 1-to-1, then $h(\hat{\theta})$ is the MLE of $h(\theta)$. **Only for MLE, not for unbiased estimators!** $\mathbb{E}[S^2] = \sigma^2$ but $\mathbb{E}[\sqrt{S^2}] \neq \sigma$.

**Example 13.5.** $\hat{\sigma}^2 = \frac{1}{n} \sum (X_i - \bar{X})^2$ is the MLE of $\sigma^2$; by invariance, $\hat{\sigma} = \sqrt{\hat{\sigma}^2}$ is the MLE of $\sigma$.

**Example 13.6** (Exponential survival function)**.** $X_i \overset{\text{iid}}{\sim} \text{Exp}(\lambda) \Rightarrow \hat{\lambda} = 1/\bar{X}$. The survival function $\bar{F}(x) = 1 - F(x) = e^{-\lambda x}$. By invariance,

$$\hat{\bar{F}}(x) \;=\; e^{-\hat{\lambda} x} \;=\; e^{-x/\bar{X}}.$$

### 13.5 Tricky case: Gamma$(\lambda, r)$

Same log-likelihood trick, but you need the *digamma* function $\psi(r) = \Gamma'(r)/\Gamma(r)$:

$$n \ln(r/\bar{X}) - n\psi(r) + \ln\big(\prod x_i\big) = 0,$$

solved numerically (Newton, bisection) for $\hat{r}$. Then $\hat{\lambda} = \hat{r}/\bar{X}$.

### 13.6 Goodness-of-fit revisited

Have we picked an appropriate distribution that models reality? Test:

$$H_0 : X_1, \dots, X_n \overset{\text{iid}}{\sim} \text{ with pmf/pdf } f(x); \qquad \alpha = P(\text{reject } H_0 \mid H_0 \text{ TRUE}) \text{ (Type I error)}.$$

Many GoF tests:

- $\chi^2$ GoF.
- Kolmogorov–Smirnov.
- Anderson–Darling.
- Shapiro–Wilk (specifically for normal).

### 13.6.1 $\chi^2$ goodness-of-fit

Partition the domain of $f(x)$ into $k$ bins. Compare observed counts to expected counts:

$$\chi_0^2 \;=\; \sum_{i=0}^{k} \frac{(O_i - E_i)^2}{E_i}.$$

If $\chi_0^2 \le \chi^2_{\alpha,k-1-s}$ (where $s = \#$ parameters estimated), fail to reject $H_0$ (accept "grudgingly"). **Use equiprobable partitions**: $E_i = n/k$. Requirements: $E_i \ge 5$ (else combine), $n \ge 30$. If $k-1-s$ is too big,

$$\chi^2_{\alpha,\nu} \approx \nu \left[ 1 - \frac{2}{9\nu} + Z_\alpha \sqrt{\frac{2}{9\nu}} \right]^3.$$

**Example 13.7** (Uniform Unif$(0,1)$, $n = 1000$, $k = 5$)**.** $E_i = 200$ for each bin. Observed: 172, 187, 200, … Compute $\chi_0^2$; degrees of freedom $\nu = k - 1 - s = 5 - 1 - 0 = 4$.

**Example 13.8** (Geometric)**.** Defects vs. frequency: $1 \to 34$, $2 \to 18$, $3 \to 7(2)$, $4 \to 9$, $5 \to 7$. Total $n = 70$. MLE $\hat{p} = 1/\bar{X} = 1/[(1)(34) + 2(18) + 3(7) + \cdots]/70 \approx 0.476$. Expected $E_x = n\,P(X = x) = n(1-\hat{p})^{x-1}\hat{p}$. Build the table; combine $E_x \ge 5$. Then $k = 4$, $s = 1$, $\nu = k - 1 - s = 2$. Compare $\chi_0^2 = 9.12$ to $\chi^2_{0.05,2} = 5.99$. **Reject: not geometric.**

**Example 13.9** (Exponential)**.** Continuous; use equiprobable bins. $1 - e^{-\lambda a_i} = i/k$, so $a_i = -\frac{1}{\lambda} \ln(1 - i/k)$. Use the MLE $\hat{\lambda} = 1/\bar{X}$, so $\hat{a}_i = -\bar{X} \ln(1 - i/k)$. Build table with $k$ bins, observed counts $O_i$, expected counts $E_i = n/k$, compute $\chi_0^2$, compare to $\chi^2_{\alpha,k-1-1}$.

**Example 13.10** (Weibull)**.** $X_i \overset{\text{iid}}{\sim} \text{Weibull}(r, \lambda)$. Two parameters, so $s = 2$. Solving for $\hat{\lambda}, \hat{r}$ is a nightmare; use Newton's method:

$$\hat{\lambda} = \left( \sum x_i^{\hat{r}} \right)^{-1/\hat{r}}, \qquad \hat{a}_i = \frac{1}{\hat{\lambda}} \left[ -\ln(1 - i/k) \right]^{1/\hat{r}}.$$

### 13.6.2 Kolmogorov–Smirnov GoF

Works especially well in low-data situations. Very conservative; need lots of bad news to reject $H_0$.

$$H_0 : X_1, \ldots, X_n \overset{\text{iid}}{\sim} \text{ from CDF } F(x).$$

Empirical CDF $\hat{F}_n(x) = \#\{X_i \le x\}/n$. Test statistic:

$$D \;=\; \max_{x \in \mathbb{R}} \left| F(x) - \hat{F}_n(x) \right|.$$

Reject if $D > D_{\alpha,n}$ (from a table specific to the distribution under test).

**Example 13.11** (Uniform, $n = 5$)**.** $X_i$: 0.039, 0.706, 0.016, 0.198, 0.793. Sort to $X_{(i)}$: 0.016, 0.039, 0.198, 0.706, 0.793. Compute $F(0) - X_{(i)} = i/n - X_{(i)}$ and $X_{(i)} - (i-1)/n$, pick the maximum of each. Combine via $D = \max(D^+, D^-) = 0.402$. Compare to $D_{0.05,5} = 0.565$. **Accept!**

**Other GoFs.** Anderson–Darling, Shapiro–Wilk (normal only), Q-Q plots (quick and dirty). "Empirical" means CDF from a sample — can fit a line through it.

### 13.7 Problem children

#### 13.7.1 Little to no data

Talk to experts: min, max, mode, quantiles, nature of observations. Try a good guess: averages/sums (Normal), bounded (Beta), success/fail (Bin/Bern), reliability (Gamma/Weibull/Lognormal), Poisson processes (e.g., particles after fallout, stock movements).

#### 13.7.2 Doesn't fit a distribution

Mix distributions, or smooth the empirical/sample distribution and bootstrap ("sample from the sample itself").

#### 13.7.3 Non-stationary

Non-homogeneous Poisson maybe; use Arena's piecewise-constant rate function.

#### 13.7.4 Multivariate / correlated

- Multivariate: e.g., height/weight. Use multivariate normal (estimating covariances is not easy).
- Serially correlated: $X_1$ affected by $X_2$ etc. Examples: unemployment, web traffic, fatigue. Use time-series models like ARMA$(p, q)$, ARTOP, EAR(1). Or just bootstrap from the empirical.

**Arena's Input Analyzer.** Small p-value? Reject $H_0$ (i.e., not a fit). $p < 0.05$ or whatever your $\alpha$ is.

## 14 Module 9: Output Analysis

*"Output is never iid. That's a lie!"*

Two main settings:

1. **Finite-horizon analysis**: terminating, replications etc. (simulation initialization matters).
2. **Steady-state analysis**: long-term.

Because input is RVs, output is RVs, so you only get estimates. Means are OK but for the full story you need variances, quantiles, probabilities. The classical statistical methods are hard to apply: there is no independence in the output stream.

### 14.1 Classical-stat warnings

Assume $Y_1, \ldots, Y_n$ are identically distributed but *not* independent.

**Mean.** $\mathbb{E}[\bar{Y_n}] = \frac{1}{n} \sum \mathbb{E}[Y_i] = \mu$. The mean is always unbiased.

**Variance.**

$$\mathrm{Var}(\bar{Y_n}) \; = \; \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} \mathrm{Cov}(Y_i, Y_j).$$

Define $R_k = \mathrm{Cov}(Y_1, Y_{1+k})$ for $k = 0, 1, 2, \ldots$ (the covariance function). Then

$$\boxed{\mathrm{Var}(\bar{Y}_n) \;=\; \frac{1}{n}\left[R_0 + 2\sum_{k=1}^{n-1}\left(1-\frac{k}{n}\right)R_k\right].}$$

(Very famous equation.) Relates the sample mean's variance to the process covariance.

**Variance parameter.**

$$\sigma_n^2 \equiv n\,\mathrm{Var}(\bar{Y}_n) \;=\; R_0 + 2\sum_{k=1}^{n-1}\left(1-\frac{k}{n}\right)R_k, \qquad \sigma^2 \equiv \lim_{n\to\infty}\sigma_n^2 \;=\; R_0 + 2\sum_{k=1}^{\infty}R_k.$$

**IID special case.** If iid, all $R_k = 0$ for $k > 0$, so $\sigma^2 = \sigma_n^2 = R_0 = \mathrm{Var}(Y_1)$.

**Dependent case.** All those covariance effects *add up.* $\sigma^2 \neq \sigma_n^2$ for large $n$; $\sigma_n^2 \gg \mathrm{Var}(Y_1)$. This wreaks havoc on CIs.

**Example 14.1** (1st-order autoregressive AR(1))**.** $Y_i = \phi Y_{i-1} + \varepsilon_i$, $|\phi| < 1$, $\varepsilon_i \sim \mathcal{N}(0, 1-\phi^2)$, $Y_0 \sim \mathcal{N}(0,1)$.

Then $R_k = \phi^{|k|}$ and

$$\sigma^2 \;=\; \sum_{k=-\infty}^{\infty}\phi^{|k|} \;=\; \frac{1+\phi}{1-\phi}.$$

With $\phi = 0.9$, you need about 19 observations to get information equivalent to one independent observation.

**Note.** $\sigma_n^2/\mathrm{Var}(Y_1)$ is "sort of" the number of observations you need to be equivalent to one independent observation.

## 14.2 Sample variance under dependence

The familiar

$$S_Y^2 \;=\; \frac{1}{n-1}\sum_{i=1}^{n}(Y_i - \bar{Y}_n)^2$$

is unbiased for $\mathrm{Var}(Y_1)$ only under iid. Under dependence,

$$\mathbb{E}[S_Y^2] \;=\; R_0 - \frac{2}{n-1}\sum_{k=1}^{n-1}\left(1-\frac{k}{n}\right)R_k \;<\; \mathrm{Var}(Y_1) \;\ll\; n\,\mathrm{Var}(\bar{Y}_n).$$

**Do not use $S_Y^2/n$ to estimate** $\mathrm{Var}(\bar{Y}_n)$**.**

## 14.3 Finite-horizon: confidence intervals

If iid normal data, the $100(1-\alpha)\%$ CI is

$$\mu \;\in\; \bar{Y}_n \pm t_{\alpha/2,n-1}\sqrt{\frac{S_Y^2}{n}}.$$

In the dependent case the true CI coverage will be much less than $(1-\alpha)$. **Be careful!**

### 14.3.1 Independent replications

Run $r$ reps of $m$ observations each; reinitialize each rep with a different random seed.

$$Z_i \;\equiv\; \frac{1}{m}\sum_{j=1}^{m} Y_{i,j}, \qquad i = 1, \dots, r.$$

“Customer $j$’s waiting time from replication $i$.” Each rep starts from the same conditions, so the $Z_i$’s are iid.

$$\bar{Z}_r = \frac{1}{r}\sum_i Z_i, \qquad S_Z^2 = \frac{1}{r-1}\sum (Z_i - \bar{Z}_r)^2.$$

$\bar{Z}_r$ is unbiased for $\mu = \mathbb{E}[\bar{Y}_m]$. Sample variance of replicate means works in the usual way; CI:

$$\mu \;\in\; \bar{Z}_r \pm t_{\alpha/2,r-1}\sqrt{\tfrac{S_Z^2}{r}}.$$

**How to shrink the CI.** Let $H = t_{\alpha/2,r-1}\sqrt{S_Z^2/r}$ be the current half-length. To achieve a smaller half-length $\epsilon$,

$$r^* \;=\; \left(\tfrac{H}{\epsilon}\right)^2 r \quad \text{(recompute all stats from } r^* \text{ reps; no guarantee)}$$

Reducing the CI by factor 10 takes 100× more replications.

### 14.3.2 Quantile CIs

Max wait time during $r$ reps. Order $W_{(1)} \le W_{(2)} \le \cdots \le W_{(r)}$. The quantile estimator

$$\hat{\xi}_p \;\equiv\; W_{(\lfloor rp + 0.5 \rfloor)}.$$

CI is bounded by

$$j = \lfloor rp + 0.5 - z_{\alpha/2}\sqrt{rp(1-p)} \rfloor, \qquad k = \lceil rp + 0.5 + z_{\alpha/2}\sqrt{rp(1-p)} \rceil.$$

**Example 14.2.** $\hat{\xi}_{0.9}$, $r = 1{,}000$: $\hat{\xi}_p = W_{(900)}$, 95% CI $= [W_{(881)}, W_{(920)}]$.

## 14.4 Steady-state simulation

Run a steady-state sim, collect $Y_1, Y_2, \dots, Y_n$. Then $\bar{Y}_n = \mu$ in the limit (good point estimator).

**Variance.** Use $\sigma^2 \equiv \lim_{n\to\infty} n \operatorname{Var}(\bar{Y}_n)$.

**Example 14.3** (MA(1))**.** $Y_{i+1} = \theta\varepsilon_i + \varepsilon_{i+1}$, $\varepsilon \sim \mathcal{N}(0,1)$. Then $\mu = \mathbb{E}[\bar{Y}_n]$, $\sigma^2 = (1+\theta)^2$.

**Example 14.4** (AR(1))**.** $Y_{i+1} = \phi Y_i + \varepsilon_{i+1}$. Then $\mu = \mathbb{E}[\bar{Y}_n]$, $\sigma^2 = (1+\phi)/(1-\phi)$.

If you can’t get $\sigma^2$ this way, use *batch means.*

## 14.5 Initializing simulations

When to start? Warmup period? Queues are not always empty/idle. How to deal with initialization bias:

- Visually (lol).
- Statistical tests on outputs.

**Welch's method.**

1. Run a lot of reps.
2. Average them (moving average).
3. Visually pick the truncation point.

Too early: bias. Too late: waste.

### 14.6 Batch means

Goal: estimate $\sigma^2$; make a CI for $\mu$.

Split $Y_1, \ldots, Y_n$ into $b$ batches of size $m$, with $n = bm$:

$$\bar{Y}_{i,m} \;=\; \frac{1}{m}\sum_{j=1}^{m} Y_{(i-1)m+j}, \quad i = 1, \ldots, b.$$

By the CLT, each batch mean is approximately normal, so

$$\bar{Y}_{1,m}, \ldots, \bar{Y}_{b,m} \overset{\text{iid}}{\approx} \mathcal{N}(\mu, \sigma^2/m).$$

**Batch means variance estimator.**

$$\hat{V}_B \;\equiv\; \frac{m}{b-1}\sum_{i=1}^{b}\left(\bar{Y}_{i,m} - \bar{Y}_n\right)^2.$$

The expectation and variance are complicated, but:

$$\mathbb{E}[\hat{V}_B] = \text{a mess}, \qquad \mathrm{Var}[\hat{V}_B] = \frac{2\sigma^2}{b-1}.$$

CI (assuming approximately symmetric):

$$\mu \;\in\; \bar{Y}_n \pm t_{\alpha/2,b-1}\sqrt{\frac{\hat{V}_B}{n}}.$$

Note $S_Z^2/m = \hat{V}_B/n$.

**Recommendation.** Use ≈ 30 batches and make the size $m$ large. This makes the independence and normality problems go away.

### 14.7 Overlapping batch means

Overlap consecutive batches: each batch shifts by 1 instead of by $m$.

$$\bar{Y}^{O}_{i,m} \;=\; \frac{1}{m}\sum_{j=i}^{i+m-1} Y_j.$$

Very correlated, but no problem!

$$\hat{V}_O \;=\; \frac{m}{n-m+1}\sum_{i}(\bar{Y}^{O}_{i,m} - \bar{Y}_n)^2.$$

Same bias as $\hat{V}_B$, less variance: $\mathrm{Var}[\hat{V}_O]/\,\mathrm{Var}[\hat{V}_B] \to 2/3$.

For large $n$, $b$ batches, dof $= \frac{3}{2}(b-1)$ with $b = n/m$.

$$\mu \;\in\; \bar{Y}_n \pm t_{\alpha/2,df}\sqrt{\tfrac{\hat{V}_O}{n}}.$$

**Use overlapping batch means for large $n$!**

### 14.8 Other variance-estimation methods

- **Spectral estimator**: uses frequency domain instead of time domain.
- **Regeneration**: simulation "restarts" at certain times → regeneration points. Break into iid blocks. But you might need to run a long time to find where the iid blocks are.
- **Standardized time series**: standardize iid RVs into a single $\mathcal{N}(\mu, \sigma^2)$. Uses Brownian-bridge processes, CLT, etc.

**Bottom line: batching makes things normal.**

## 15 Module 10: Comparing Systems

How to analyze a system and compare them:

1. 1 system: CIs from baby stats.
2. 2 systems: paired-system CIs.
3. ≥ 2 systems: ranking and selection.

### 15.1 CI for $\mu$ of iid normal data

**One sample.** $X_1, \dots, X_n \sim \mathcal{N}(\mu, \sigma^2)$, $\mu$ unknown, $\sigma^2$ unknown.

$$\bar{X}_n = \tfrac{1}{n}\sum X_i \sim \mathcal{N}(\mu, \sigma^2/n), \qquad S_X^2 = \tfrac{1}{n-1}\sum (X_i - \bar{X}_n)^2 \sim \sigma^2 \chi^2_{n-1}/(n-1).$$

Then $T = (\bar{X}_n - \mu)/\sqrt{S_X^2/n} \sim t(n-1)$. So

$$1 - \alpha = P(-t_{\alpha/2,n-1} \le T \le +t_{\alpha/2,n-1}),$$

giving the CI

$$\boxed{\mu \;\in\; \bar{X}_n \pm t_{\alpha/2,n-1}\sqrt{S_X^2/n}.}$$

**Two systems: difference of means.** Take $X_i \overset{\text{iid}}{\sim} \mathcal{N}(\mu_X, \sigma_X^2)$ and $Y_i \overset{\text{iid}}{\sim} \mathcal{N}(\mu_Y, \sigma_Y^2)$; want a CI for $\mu_X - \mu_Y$. Three flavors:

1. Pooled CI: $\sigma_X^2 = \sigma_Y^2$, both unknown.
2. Approximate CI: $\sigma_X^2 \neq \sigma_Y^2$, both unknown.
3. Paired CI: when $\mathrm{Cov}(X_i, Y_i) > 0$ (collecting in pairs).

**Note:** if the CI contains zero, the difference is inconclusive. The position of zero (left or right of the interval) tells you which mean is larger.

### 15.1.1 (1) Pooled CI

When $\sigma_X^2 = \sigma_Y^2$ (equal but unknown), use the pooled variance:

$$S_p^2 \;\equiv\; \frac{(n-1)S_X^2 + (m-1)S_Y^2}{n+m-2},$$

and

$$\mu_X - \mu_Y \;\in\; \bar{X}_n - \bar{Y}_m \pm t_{\alpha/2,\nu} S_p \sqrt{\frac{1}{n} + \frac{1}{m}}, \quad \nu = n+m-2.$$

### 15.1.2 (2) Approximate CI

When $\sigma_X^2 \neq \sigma_Y^2$, both unknown:

$$\mu_X - \mu_Y \;\in\; \bar{X} - \bar{Y} \pm t_{\alpha/2,\nu} \sqrt{\tfrac{S_X^2}{n} + \tfrac{S_Y^2}{m}}$$

with approximate degrees of freedom (round down)

$$\nu \;\equiv\; \frac{(S_X^2/n + S_Y^2/m)^2}{(S_X^2/n)^2/(n+1) + (S_Y^2/m)^2/(m+1)} - 2.$$

### 15.1.3 (3) Paired CI

Inter-pair iid, intra-pair not iid. Eliminates extraneous noise; use when possible.

E.g., the same person parks two different cars. Take $D_i = X_i - Y_i \overset{\text{iid}}{\sim} \mathcal{N}(\mu_D, \sigma_D^2)$ with

$$\sigma_D^2 \;=\; \sigma_X^2 + \sigma_Y^2 - 2\,\mathrm{Cov}(X_i, Y_i).$$

We hope $\mathrm{Cov}(X_i, Y_i)$ is large (and positive!) so $\sigma_D^2$ is reduced.

$$\bar{D} = \tfrac{1}{n}\sum D_i \sim \mathcal{N}(\mu_D, \sigma_D^2/n), \qquad S_D^2 = \tfrac{1}{n-1}\sum (D_i - \bar{D})^2,$$

$$\mu_D \;\in\; \bar{D} \pm t_{\alpha/2,n-1}\sqrt{S_D^2/n}.$$

## 15.2 Indifference-zone approach

For $k \geq 2$ systems. Specify the probability of picking the best system over its competitors. Other techniques (CI, ANOVA) tell you "at least one is better than the rest" — not enough.

**Selection criteria.**

- Greatest mean? Normal data.
- Highest probability? Binomial / Bernoulli.
- Best player overall? Multinomial.

**Setup.** $k$ normal populations $\mathcal{N}(\mu_i, \sigma_i^2)$. Order: $\mu_{[1]} \leq \mu_{[2]} \leq \cdots \leq \mu_{[k]}$. We want the largest mean but $\mu_{[k]}$ is unknown.

**Correct selection.** Pre-choose $\delta^* > 0$ (smallest difference worth detecting) and $P^* \in (1/k, 1)$:

$$P(CS) \;\geq\; P^* \text{ whenever } \mu_{[k]} - \mu_{[k-1]} \geq \delta^*.$$

Detection probability depends on $n$, $\sigma^2$, and the true means $\mu_i$. If $\mu_{[k]} - \mu_{[k-1]} \geq \delta^*$ you are in the *preference zone*; else in the *indifference zone.*

**If you pick $\mu_{[k-1]}$ by mistake.** Not a big deal — they were so close anyway.

**Procedures.** Many ways:

- Bechhofer (single-stage).
- Rinott (two-stage).
- Kim & Nelson (sequential).

### 15.2.1 Bechhofer single-stage (NORMAL MEANS)

Make a decision at once. **Requires known common variance.** For $k$ competitors, $P^*$, $\delta^*/\sigma$, look up $n$ from a table (or get from a multivariate normal quantile, or from a simulation). Sample $n$ observations from each of the $k$ competitors. Take sample means $\bar{\mu}_k$; pick the largest.

**Slippage configuration.** $\mu_{[1]} = \mu_{[2]} = \cdots = \mu_{[k-1]} = \mu_{[k]} - \delta^*$. $\mu_{[k]}$ is bigger than all the rest by a fixed amount.

**Least-favorable configuration.** All others minimized to be as close as possible → $P(CS)$ is minimized! Very hard to achieve.

If $\delta^*$ is large and $P^*$ is small, you need smaller $n$.

### 15.2.2 Bechhofer-Sobel-Huyett, etc.

**(b) Pick with highest probability (or other parameter).** Goal: select Bernoulli population with highest success parameter. Most of $X$, best of $Y$, etc. (**Spiritually the same as (a), Bechhofer.**) OK if independent.

$k$ Bernoulli populations with $P_1, \ldots, P_k$, ordered $P_{[1]} \leq \cdots \leq P_{[k]}$. Take $(P^*, \Delta^*)$:

$$\tfrac{1}{k} < P^* < 1, \qquad 0 < \Delta^* < 1.$$

$P(CS) \geq P^*$ whenever $P_{[k]} - P_{[k-1]} \geq \Delta^*$ ("when the difference is pretty big, don't fuck up selecting $P_{[k]}$; want to pick it with probability $P^*$").

**Example 15.1.** $k = 4$ treatments, $\Delta^* = 0.10$, $P^* = 0.95$. Table says $n = 212$. Pick $Y_{k,n}$ with the most successes.

### 15.2.3 (c) Multinomial

"Best candidate?" "Most-watched show?" "Which configuration maximizes throughput?"

Different from Bernoulli because candidates compete with each other. This generalizes Bernoulli/Binomial. Think of it like binning.

- $k$ outcomes.
- $p_i$ = prob of category $i$.
- $n$ independent replications.
- $Y_i$ = # outcomes in category $i$.

$Y = (Y_1, \ldots, Y_k)$ is a $k$-variate discrete vector RV. If $P\{Y_1 = y_1, \ldots, Y_k = y_k\} = \frac{n!}{\prod y_i!} \prod p_i^{y_i}$ with each $p_i > 0$ summing to 1, then $Y$ is multinomial.

**Example 15.2.** Red, Blue, Green with $p_R = 3/6, p_B = 2/6, p_G = 1/6$. Toss 5 times, find $P\{Y = (3, 0, 2)\}$:

$$P = \frac{5!}{3!\,0!\,2!}(3/6)^3(2/6)^0(1/6)^2.$$

**Example 15.3** (Selection)**.** Same setup but don't know $p_i$ — which candidate do you like? Run simulation (say 5 reps), use the multinomial formula to calculate probabilities of R, B, G being picked. Randomize ties.

**Selection probability.** $X_j = (X_{1j}, \ldots, X_{kj})$, $X_{kj} = 1$ if person $j$ picked category $k$. Probs $p_1, \ldots, p_k$ unknown; $P_{[k]}$ is the "best."

GOAL: Pick the category with the largest $P_{[k]}$. NOTE: This is an extension to categorical RVs!

**Indifference zone.** $\frac{1}{k} < P^* < 1$, $P_{[k]}/P_{[k-1]} \geq \theta^*$ ("ratio is at least $\theta$"). Procedure achieves $P(CS) \geq P^*$ in least-favorable config:

$$P_{[1]} = P_{[2]} = \cdots = P_{[k-1]} = \frac{1}{\theta^* + k - 1}, \qquad P_{[k]} = \frac{\theta^*}{\theta^* + k - 1}.$$

Very hard to achieve! (Bechhofer-Elmaghraby-Morse.)

**Example 15.4.** $k = 3$ kinds of colas, need $P(CS) \geq 0.95$, "ratio at least 1.4" ($\theta^*$). Table: $n = 186$. Say $Y_{1,186} = 26, Y_{2,186} = 110, Y_{3,186} = 26$. Winner: cola #2.

**Curtailed sampling.** Stop when cola #2 cannot win.

**Sequential.** Sample one at a time, stop when there is a clear winner.

# 16 Module 9 (continued): Variance Reduction and Comparing Sims

## 16.1 Three approaches to comparing simulations

1. Confidence intervals.
2. Variance reduction.
3. Ranking & selection.

### 16.2 (1) Confidence intervals (recap)

Run replications with different PRNG seeds. Compare approximate CIs using baby statistics. You can also use paired CIs. $Z_{i,j}$ = data point $j$ from replication $i$.

### 16.3 (2) Variance reduction

#### 16.3.1 Common Random Numbers (CRN)

Use the same PRNGs to make decisions about each system. E.g., System A has service time $\hat{\theta}_A$, System B has $\hat{\theta}_B$. Estimate $\hat{\theta}_A \to \theta_A$, $\hat{\theta}_B \to \theta_B$. OK if $\hat{\theta}_A < \hat{\theta}_B$ (maybe), but

$$\mathrm{Var}(\hat{\theta}_A - \hat{\theta}_B) \; = \; \mathrm{Var}(\hat{\theta}_A) + \mathrm{Var}(\hat{\theta}_B) \quad \text{(when independent)}.$$

So we lose conviction. But with the same PRNGs,

$$\mathrm{Var}(\hat{\theta}_A - \hat{\theta}_B) \; = \; \mathrm{Var}(\hat{\theta}_A) + \mathrm{Var}(\hat{\theta}_B) - 2\ \mathrm{Cov}(\hat{\theta}_A, \hat{\theta}_B).$$

Adds a lot of covariance! So you can drive down $\mathrm{Var}(\hat{\theta}_A - \hat{\theta}_B) \to$ **tighter CIs**. The covariance is intentional.

**Example 16.1.** Use $U = 0.17, 0.23, 0.46, 0.76, \ldots$ both times: run $\hat{\theta}_A$ with these, run $\hat{\theta}_B$ with these too.

#### 16.3.2 Antithetic PRNGs

This time, introduce *negative* correlation. "Opposite day."

$$\mathrm{Var}\Big(\tfrac{\hat{\theta}_A + \hat{\theta}_B}{2}\Big) = \tfrac{1}{4}\Big[\mathrm{Var}(\hat{\theta}_A) + \mathrm{Var}(\hat{\theta}_B) + 2\ \mathrm{Cov}(\hat{\theta}_A, \hat{\theta}_B)\Big] \; < \; \mathrm{Var}(\hat{\theta}_A)/2.$$

Use $U = 0.17, 0.23, \ldots$ for $\hat{\theta}_A$, then $1 - U = 0.83, 0.77, \ldots$ for $\hat{\theta}_B$; average the two estimators.

**The whole goal of these two methods is variance reduction.**

### 16.4 (3) Ranking and selection (recap)

(Covered in detail in Module 10 above.) Specify $P(CS) \geq P^*$, $\delta^*$. Many techniques:

- Bechhofer (single-stage).
- Rinott (two-stage).
- Kim & Nelson (sequential).

## 17 Closing Notes

A few miscellaneous identities and tricks scattered throughout the material:

**Linearity of expectation.** $\mathbb{E}[X + Y] = \mathbb{E}[X] + \mathbb{E}[Y]$ always.

**Variance with correlation.** $\mathrm{Var}(X \pm Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) \pm 2\ \mathrm{Cov}(X, Y)$.

**Conditional probability vs. independence.**

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)}; \qquad \text{independent} \implies P(A \mid B) = P(A).$$

**Sample mean distribution.** $X_i \overset{\text{iid}}{\sim} \mathcal{N}(\mu, \sigma^2) \implies \bar{X} \sim \mathcal{N}(\mu, \sigma^2/n)$. The CLT extends this approximately to any distribution with finite variance.

**Laplace**$(\mu, b)$**.** $f(x) = \frac{1}{2b} \exp(-|x-\mu|/b)$. $\mathbb{E} = \mu =$ median $=$ mode. Var $= 2b^2$. MGF $= e^{\mu t}/(1-b^2t^2)$ for $|t| < 1/b$.

**Inverse-transform summary.**

- Exponential: $X = -\frac{1}{\lambda} \ln(U)$.
- Weibull: $X = -\frac{1}{\lambda}[\ln(1-U)]^{1/\beta}$.
- Geometric: $X = \lceil \ln(U)/\ln(1-p) \rceil$, or count Bernoulli trials.
- Triangle: $X = \sqrt{2U}$ for $U < 1/2$, else $2 - \sqrt{2(1-U)}$.
- Normal: closed form not available; use table lookup, or Box-Muller, or polar.

**Crude normal approximation.**

$$Z = \Phi^{-1}(U) \;\approx\; \frac{U^{0.135} - (1-U)^{0.135}}{0.1975}.$$

**Standardization.** $X \sim \mathcal{N}(\mu, \sigma^2)$, $Z = (X-\mu)/\sigma \sim \mathcal{N}(0,1)$. To generate $\mathcal{N}(3, 16)$: $X = 3 + 4Z = 3 + 4\Phi^{-1}(U)$.