Sample Measures

Sample Mean: $\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$
Sample Variance: $s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(y_i - \bar{y})^2$
Sample Variance (shortcut): $s^2 = \frac{1}{n-1}\left[\sum_{i=1}^{n} y_i^2 - n\bar{y}^2\right]$

Basic Probability

$A \cup B$: A "or" B (union)
$A \cap B$: A "and" B (intersection)
$\bar{A}$: "not" A (complement)
Distributive Laws:

$$\begin{aligned} A \cap (B \cup C) &= (A \cap B) \cup (A \cap C) \\ A \cup (B \cap C) &= (A \cup B) \cap (A \cup C) \end{aligned}$$

De Morgan's Laws:

$$\overline{(A \cap B)} = \bar{A} \cup \bar{B}, \quad \overline{(A \cup B)} = \bar{A} \cap \bar{B}$$

Sample Space $\mathcal{S}$: set of all possible experimental outcomes.
Axioms: $P(\mathcal{S}) = 1$, $0 \leq P(A) \leq 1 \quad \forall\ A \in \mathcal{S}$, and if $\{A_1, A_2, \cdots, A_n\}$ are pairwise disjoint, $P(A_1 \cup \cdots \cup A_n) = \sum_{i=1}^{n} P(A_i)$
$A|B$: A "given" B (condition)
$P(A|B) = P(A \cap B)/P(B)$ (conditional probability)
$P(A|B) = P(A)$ iff A and B are independent
$P(A \cap B) = P(A)P(B)$ iff A and B are independent
$P(A \cap B) = P(A|B)P(B)$ (multiplicative rule)
$P(A \cup B) = P(A) + P(B) - P(A \cap B)$ (additive rule)
$P(A) = 1 - P(\bar{A})$
$P(A) = \sum_{i=1}^{n} P(A|B_i)P(B_i)$ iff $B_i's$ disjoint/exhaustive (law of total probability)
$P(A|B) = P(B|A)P(A)/P(B)$ (Bayes' rule)

Counting Tools

mn rule: by selecting one outcome from $a_1, \cdots, a_m$ and one from $b_1, \cdots, b_n$, there are mn possible outcomes overall
If each experiment has the same number of outcomes, then the total number of possible outcomes is $(\#\text{outcomes})^{(\#\text{experiments})}$
Number of ways to permute n objects taken r at a time: $P_r^n = n!/(n-r)!$
Number of ways to combine n objects taken r at a time: $C_r^n = n!/[r!(n-r)!]$

Probability Distributions: Discrete

Probability mass function (pmf): $p(y) = P(Y = y)$
Cumulative distribution function (cdf):

$$F(y) = \sum_{y'=-\infty}^{y} p(y'), \quad p(y) = F(y) - F(y-1)$$

$0 \leq p(y) \leq 1$; $\sum_{y=-\infty}^{\infty} p(y) = 1$; $P(Y = y) = p(y)$
$P(a \leq Y \leq b) = \sum_{y=a}^{b} p(y)$

Probability Distributions: Continuous

Probability density function (pdf): $f(y)$
Cumulative distribution function (cdf):

$$F(y) = \int_{y'=-\infty}^{y} f(y')dy', \quad f(y) = \frac{d}{dy}F(y)$$

$f(y) \geq 0$; $\int_{-\infty}^{\infty} f(y)dy = 1$; $P(Y = y) = 0$
$P(a \leq Y \leq b) = \int_{y=a}^{b} f(y)dy = F(b) - F(a)$

Expectation and Variance

Expected Value (Population Mean):

$$E[Y] = \mu = \sum_{-\infty}^{\infty} yp(y) = \int_{-\infty}^{\infty} yf(y)dy$$

Law of the Unconscious Statistician:

$$E[g(Y)] = \sum_{-\infty}^{\infty} g(y)p(y) = \int_{-\infty}^{\infty} g(y)f(y)dy$$

Variance:

$$V[Y] = \sigma^2 = \sum_{-\infty}^{\infty}(y - E[Y])^2 p(y) = \int_{-\infty}^{\infty}(y - E[Y])^2 f(y)dy$$

Shortcut: $V[Y] = E[Y^2] - (E[Y])^2$
Standard Deviation: $\sigma = \sqrt{V[Y]}$

Bivariate Distributions

Conditional PMF: $p(y_1|y_2) = p(y_1, y_2)/p_2(y_2)$ (and v-v)
Conditional PDF: $f(y_1|y_2) = f(y_1, y_2)/f_2(y_2)$ (and v-v)
Conditional CDF: $F(y_1|y_2) = P(Y_1 \leq y_1|Y_2 = y_2)$ (and v-v)
Y_1 and Y_2 are independent if, e.g., $F(y_1, y_2) = F(y_1)F(y_2)$, $p(y_1, y_2) = p(y_1)p(y_2)$, $f(y_1, y_2) = f(y_1)f(y_2)$, or if you can decompose $f(y_1, y_2) = g(y_1)h(y_2)$.
$E[g(Y_1, Y_2)] = \sum_{y_1=-\infty}^{\infty}\sum_{y_2=-\infty}^{\infty} g(y_1, y_2)p(y_1, y_2)$ (discrete)
$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} dy_1 dy_2 g(y_1, y_2)f(y_1, y_2)$ (continuous)
$E[g(Y_1)h(Y_2)] = E[g(Y_1)]E[h(Y_2)]$ if Y_1 and Y_2 independent

Covariance and Correlation

$\text{Cov}(Y_1, Y_2) = E[(Y_1 - \mu_1)(Y_2 - \mu_2)] = E[Y_1Y_2] - E[Y_1]E[Y_2]$
$\text{Corr}(Y_1, Y_2) = \rho = \text{Cov}(Y_1, Y_2)/(\sigma_1\sigma_2)$; $-1 \leq \rho \leq 1$
If Y_1 and Y_2 are independent, then $\text{Cov}(Y_1, Y_2) = 0$.

Linear Functions of R.V.'s

If $U_1 = \sum_{i=1}^{n} a_i Y_i$ and $U_2 = \sum_{j=1}^{m} b_j X_j$, then:
(a) $E[U_1] = \sum_{i=1}^{n} a_i E[Y_i]$,
(b) $V[U_1] = \sum_{i=1}^{n} a_i^2 V[Y_i] + 2\sum\sum_{1 \leq i < j \leq n} a_i a_j \text{Cov}(Y_i, Y_j)$ (with double sum over all pairs (i, j) s.t. $i < j$), and
(c) $\text{Cov}(U_1, U_2) = \sum_{i=1}^{n}\sum_{j=1}^{m} a_i b_j \text{Cov}(Y_i, X_j)$

Conditional Expectation/Variance

$E[g(Y_1)|Y_2 = y_2] = \sum_{-\infty}^{\infty} g(y_1)p(y_1|y_2) = \int_{-\infty}^{\infty} g(y_1)f(y_1|y_2)$
Unconditional: $E[Y_1] = E[E[Y_1|Y_2]]$
$V[Y_1|Y_2 = y_2] = E[Y_1^2|Y_2 = y_2] - (E[Y_1|Y_2 = y_2])^2$
Unconditional: $V[Y_1] = E[V[Y_1|Y_2]] + V[E[Y_1|Y_2]]$

Method of Distribution Functions

(a) Find the region $U = u$ in the $(y_1, \cdots, y_n)$ space
(b) Find the region $U \leq u$
(c) Find $F_U(u) = P(U \leq u)$ by integrating $f(y_1, \cdots, y_n)$ over the region $U \leq u$
(d) Set $f_U(u) = dF_U(u)/du$

Method of Transformations (for monotone functions)

(a) Find the inverse function $y = h^{-1}(u)$
(b) Evaluate dh^{-1}/du
(c) Set $f_U(u) = f_Y[h^{-1}(u)]|dh^{-1}/du|$

Moment Generating Functions

$m_Y(t) = E[e^{tY}] = \sum_{-\infty}^{\infty} e^{ty} p(y) = \int_{-\infty}^{\infty} e^{ty} f(y) dy$
$m_Y(t) = 1 + t\mu_1' + (t^2/2!)\mu_2' + \cdots$
$\mu_k' = \sum_{-\infty}^{\infty} y^k p(y) = \int_{-\infty}^{\infty} y^k f(y) dy = (d^k m/dt^k)|_{t=0}$
$E[Y] = \mu_1'$ and $E[Y^2] = V[Y] + (E[Y])^2 = \mu_2'$
$E[e^{tg(Y)}] = \sum_{-\infty}^{\infty} e^{tg(y)} p(y) = \int_{-\infty}^{\infty} e^{tg(y)} f(y) dy$
If $X = a_1 Y_1 + a_2 Y_2 + b$, $m_X(t) = e^{bt} m_{Y_1}(a_1 t) m_{Y_2}(a_2 t)$
If $Y_1, Y_2, ..., Y_n$ are independent random variables, and $Y = a_1 Y_1 + a_2 Y_2 + ... + a_n Y_n$, then $m_Y(t) = m_{Y_1}(a_1 t) \cdot m_{Y_2}(a_2 t) \cdot \cdots \cdot m_{Y_n}(a_n t)$.
The mgf, if it exists, uniquely determines the probability distribution.

Order Statistics

Place iid data $Y_1, \cdots, Y_n$ into ascending order: $Y_{(1)}, \cdots, Y_{(n)}$. The pdf for $Y_{(k)}$ is

$$g_{(k)}(y) = \frac{n!}{(k-1)!(n-k)!}[F(y)]^{k-1}[1-F(y)]^{n-k} f(y)$$

Sampling Dists Related to Normal Dist

Assume $Y_1, \cdots, Y_n$ are iid samples from $\mathrm{N}(\mu, \sigma^2)$. Then
$\bar{Y} = (1/n)\sum_{i=1}^n Y_i \sim \mathrm{N}(\mu, \sigma^2/n)$
$\sum_{i=1}^n Z_i^2 = \sum_{i=1}^n ((Y_i - \mu)/\sigma)^2 \sim \chi^2(n)$
$(n-1)S^2/\sigma^2 = (1/\sigma^2)\sum_{i=1}^n (Y_i - \bar{Y})^2 \sim \chi^2(n-1)$
$(\bar{Y} - \mu)/(S/\sqrt{n}) \sim \mathrm{t}(n-1)$
If Z is standard normal and W is χ^2-distributed with ν dof, and Z and W are independent r.v.'s, then $T = Z/\sqrt{W/\nu}$ is t-distributed with ν dof (degrees of freedom)
If $W_1 \sim \chi^2(\nu_1)$ and $W_2 \sim \chi^2(\nu_2)$, and W_1 and W_2 are independent r.v.'s, then $F = (W_1/\nu_1)/(W_2/\nu_2)$ is F-distributed with ν_1 numerator dof and ν_2 denominator dof

Central Limit Theorem

Let $Y_1, \cdots, Y_n$ be iid draws from an arbitrary distribution with known μ and σ^2 ($< \infty$). If $U_n = \sqrt{n}(\bar{Y} - \mu)/\sigma$, then

$$\lim_{n\to\infty} P(U_n \leq u) = \Phi(u), \ \forall\ u\,,$$

i.e., one may assume $\bar{Y} \sim \mathrm{N}(\mu, \sigma^2/n)$. RoT: $n \gtrsim 30$.

Descriptive Rules

Empirical Rule: if the distribution is approximately mound-shaped and symmetric, then $\approx$ 68%, 95%, and all of the measurements will lie in the ranges $\mu \pm \sigma$, $\mu \pm 2\sigma$, and $\mu \pm 3\sigma$, respectively.
Tchebysheff's Theorem: the probability content within the range $\mu \pm k\sigma$ is at least $1 - 1/k^2$.

Generally Helpful Tidbits

$E[Y^a] = \mu_a' = \sum_{-\infty}^{\infty} y^a p(y)$ or $\int_{-\infty}^{\infty} y^a f(y) dy$
$E[Y] = \mu_1' = \mu$
$V[Y] = \mu_2 = \sigma^2 = \sum_{-\infty}^{\infty} (y - E[Y])^2 p(y)$ or $\int_{-\infty}^{\infty} (y - E[Y])^2 f(y) dy$
$V[Y] = E[Y^2] - (E[Y])^2$
$E[Y(Y-1)] = E[Y^2] - E[Y] = V[Y] + (E[Y])^2 - E[Y]$
$\sigma = \sqrt{V[Y]}$

Often-Used Distributions

Binomial Distribution – $Y \sim \mathrm{Bin}(n,p)$

$p(y) = \frac{n!}{y!(n-y)!} p^y (1-p)^{n-y}$, $y \in \{0, \mathbb{N}\}$
$E[Y] = np$; $V[Y] = np(1-p)$; $m_Y(t) = [pe^t + (1-p)]^n$
If $n \gtrsim 100$, $p \lesssim 0.01$, and $np \lesssim 7$, approximate the binomial with the Poisson, i.e., $p(y; n, p) \to p(y; \lambda = np)$

Geometric Distribution – $Y \sim \mathrm{Geom}(p)$

$p(y) = (1-p)^{y-1} p$, $y \in \mathbb{N}$
$E[Y] = 1/p$; $V[Y] = (1-p)/p^2$; $m_Y(t) = pe^t/[1 - (1-p)e^t]$

Negative Binomial Distribution – $Y \sim \mathrm{NB}(r,p)$

$p(y) = \binom{y-1}{r-1} p^r (1-p)^{y-r}$, $y \in [r, r+1, \cdots]$
$E[Y] = r/p$; $V[Y] = r(1-p)/p^2$; $m_Y(t) = (pe^t/[1-(1-p)e^t])^r$

Hypergeometric Dist – $Y \sim \mathrm{Hypergeometic}(r, n, N)$

$p(y) = C_y^r C_{n-y}^{N-r}/C_n^N = \binom{r}{y}\binom{N-r}{n-y}/\binom{N}{n}$, $0 \leq y \leq n$, subject to $y \leq r$ and $n - y \leq N - r$
$E[Y] = nr/N$; $V[Y] = n(r/N)[(N-r)/N][(N-n)/(N-1)]$

Poisson Distribution – $Y \sim \mathrm{Pois}(\lambda)$

$p(y) = \frac{\lambda^y}{y!} e^{-\lambda}$, $y \in \{0, \mathbb{N}\}$
$E[Y] = V[Y] = \lambda$; $m_Y(t) = \exp[\lambda(e^t - 1)]$

Uniform Distribution – $Y \sim \mathrm{Uniform}(a,b)$

$f(y) = 1/(b-a)$, $a \leq y \leq b$
$E[Y] = (a+b)/2$; $V[Y] = (b-a)^2/12$;
$m_Y(t) = (e^{bt} - e^{at})/[t(b-a)]$

Normal Distribution – $Y \sim \mathrm{N}(\mu, \sigma^2)$

$f(y) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(y-\mu)^2/2\sigma^2} \quad -\infty < y < \infty$
$E[Y] = \mu$; $V[Y] = \sigma^2$; $m_Y(t) = \exp(\mu t + t^2\sigma^2/2)$
If $Y \sim N(\mu, \sigma)$, then $Z = (Y - \mu)/\sigma$; $Z \sim \mathrm{N}(0,1)$.
$P(Y \leq y) = \Phi\left(\frac{y-\mu}{\sigma}\right) = \Phi(z)$ (non-analytic function)

Gamma Distribution – $Y \sim \mathrm{Gamma}(\alpha, \beta)$

$f(y) = y^{\alpha-1} e^{-y/\beta}/[\beta^\alpha \Gamma(\alpha)]$, $0 \leq y < \infty$, $\alpha, \beta > 0$
$\Gamma(\alpha) = \int_0^\infty y^{\alpha-1} e^{-y} dy = (\alpha-1)\Gamma(\alpha-1)$
$\Gamma(\alpha) = (\alpha-1)!$, for $\alpha \in \mathbb{N}$.
$E[Y] = \alpha\beta$; $V[Y] = \alpha\beta^2$; $m_Y(t) = (1 - \beta t)^{-\alpha}$
$\alpha = 1 \Rightarrow$ exponential distribution – $Y \sim \mathrm{Exp}(\beta)$, pdf: $f(y) = \frac{1}{\beta} e^{-y/\beta}$, CDF: $F(y) = 1 - e^{-y/\beta}$
$\beta = 2$, $\alpha = \nu/2$, $\nu \in \mathbb{N} \Rightarrow$ chi-square distribution, $\chi^2(\nu)$.
The sum of squares of n independent standard normal has a $\chi^2(n)$ distribution.

Beta Distribution – $Y \sim \mathrm{Beta}(\alpha, \beta)$

$f(y) = y^{\alpha-1}(1-y)^{\beta-1}/B(\alpha, \beta)$, $0 \leq y \leq 1$
$B(\alpha, \beta) = \Gamma(\alpha)\Gamma(\beta)/\Gamma(\alpha+\beta)$
$E[Y] = \alpha/(\alpha+\beta)$; $V[Y] = \alpha\beta/[(\alpha+\beta)^2(\alpha+\beta+1)]$