Probability Theory

$$P(A \cup B) = P(A) + P(B) \text{ disjoint}$$
$$P(B \cap A^c) = P(B) - P(A \cap B)$$
$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$
$$\text{If } A \subset B, P(A) \leq P(B)$$
$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$
$$P(A \cap B) = P(A)P(B) \text{ indep.}$$
$$\text{Mutually exclusive } \implies P(A \cap B) = \emptyset$$

Counting

w/o rep. and ordered: $n!/(n-r)!$

w/ rep. and ordered: n^r

w/o rep. and unordered: n choose r

w/ rep. and unordered: $\binom{n+r-1}{r}$

cdf, pdf, pmf

$$\text{cdf: } \lim_{x \to -\infty} F(x) = 0, \lim_{x \to \infty} F(x) = 1$$
$F_X(x)$ is nondecreasing, derivative is > 0

$F_X(x)$ is right-continuous: $\lim_{x \downarrow x_0} F(x) = F(X_0)$
$$P(X \leq x) = F_X(x) = \int_{-\infty}^{x} f_X(t)dt$$
$$\frac{d}{dx}F_X(x) = f_X(x)$$
$$P(a < X < b) = F_X(b) - F_X(b)$$
$$\text{pdf, pmf: } f_X(x) \geq 0 \quad \forall x$$
$$\sum_x f_X(x) = 1 \text{ or } \int_{-\infty}^{\infty} f_X(x)dx = 1$$

Transformations

*X had pdf f and Y = g(X), where g is monotone.

f(x) is continuous and $g^{-1}(y) = x$ has cont. deriv.

If $g(X) = 1/X$ let $y = g(x) \implies g^{-1}(y) = 1/y$
$$f_Y(y) = f_X(g^{-1}(y))|\frac{d}{dy}g^{-1}(y)|$$

*Suppose there exists a partition, $A_0, A_1, \cdots, A_k$ s.t. $P(X \in A_0) = 0$ and f is continuous on each A

$g(x) = g_i(x)$ for $x \in A_i, g_i(x) is monotone on A_i$

$g_i^{-1}(y)$ has continuous derivative. Then
$$f_Y(y) = \sum_{i=1}^{k} f_X(g_i^{-1}(y))|\frac{d}{dy}g_i^{-1}(y)|$$

Expected Values

$$Eg(X) = \int_{-\infty}^{\infty} g(x)f_X(x)dx \text{ if cont.}$$
$$Eg(X) = \sum_x g(x)P(X = x) \text{ if discr.}$$
$$\text{Identity: } x\binom{n}{x} = n\binom{n-1}{x-1}$$
$$E(ag_1(X) + bg_2(X) + c = aEg_1(X) + bEg_2(X) + c$$
$$\text{If } g_1(x) \geq 0, Eg_1(X) \geq 0$$
$$\text{If } g_1(x) \geq g_2(x), Eg_1(x) \geq Eg_2(x)$$
$$\text{If } a \geq g_1(x) \leq b, a \geq Eg_1(x) \leq b$$
$$EX = \sum_x xP(X = x)$$
$$E[XY] = \sum_x \sum_y xyf(x, y)$$

Moments and mgfs

$$Var(aX + b) = a^2 VarX$$
$$VarX = E(X - EX)^2 = EX^2 - (EX)^2$$
$$\text{Identity: } x^2\binom{n}{x} = xn\binom{n-1}{x-1}$$
$$M_X(t) = Ee^{tX} = \int_{-\infty}^{\infty} e^{tx} f_X(x)dx \text{ cont.}$$
$$M_X(t) = Ee^{tX} = \sum_x e^{tx} P(X = x) \text{ discr.}$$
$$EX^n = M_X^{(n)}(0) = \frac{d^n}{dt^n} M_X(t)|_{t=0}$$
$$\text{Binomial formula: } \sum_{x=0}^{n} \binom{n}{x} u^x v^{n-x} = (u + v)^n$$

Convergence of mgfs implies convergence of cdfs
$$\exp(x) = \lim_{n \to \infty} \left(1 + \frac{x}{n}\right)^n$$
$$\exp(x) = \sum_{n=0}^{\infty} \frac{x^n}{x!}$$
$$M_{aX+b} = e^{bt} M_X(at)$$
$$\frac{d}{d\theta} \int_a^b f(x, \theta) = \int_a^b \frac{\partial}{\partial\theta} f(x, \theta)$$
$$log(1 - x) = -\sum_{k=1}^{\infty} \frac{x^k}{k}$$

Var X DNE if the 2nd moment is infinite

Exponential Families

$$f(x|\vec{\theta}) = h(x)c(\vec{\theta}) \exp\left(\sum_{i=1}^{k} w_i(\vec{\theta})t_i(x)\right)$$
$$f(x|\eta) = h(x)c^*(\eta) \exp\left(\sum_{i=1}^{k} \eta_i t_i(x)\right)$$
$$E(t_j(X)) = -\frac{\partial}{\partial\eta_j} \log c^*(\eta)$$
$$Var(t_j(X)) = -\frac{\partial^2}{\partial\eta_j^2} \log c^*(\eta)$$

Inequalities and Identities

$$P(g(X) \geq r) \leq \frac{Eg(X)}{r}$$

Joint and Marginal Distributions

Refer to properties of pdfs/pmfs
$$f_X(x) = \sum_y f_{X,Y}(x, y)$$
$$f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x, y)dx$$
$$Eg(X, Y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x, y) f_{X,Y}(x, y)dxdy$$
$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{X,Y}(x, y)dxdy = 1$$

Conditional Distributions + Indep.

$$f(y|x) = P(Y = y|X = x) = \frac{f(x, y)}{f_X(x)}$$
$$E(g(Y)|x) = \sum_y g(y)f(y|x) \text{ disc.}$$
$$E(g(Y)|x) = \int_y g(y)f(y|x)dy \text{ cont.}$$
$$E[E[Y|X]] = \int_y E[Y|x]f_X(x)dx \text{ cont.}$$
$$\text{If Indep. } f(x, y) = f_X(x)f_Y(y)$$
$$\text{Indep. iff } f(x, y) = g(x)h(y)$$
$$\text{If indep. } E(g(X)h(Y)) = (Eg(X))(Eh(Y))$$

Covariance and Correlation

$$EX = E(E(X|Y))$$
$$Cov(X, Y) = E((X - \mu_X)(Y - \mu_Y))$$
$$Cov(X, Y) = EXY - \mu_X\mu_Y$$
$$\rho_{XY} = \frac{Cov(X, Y)}{\sigma_X\sigma_Y}$$
$$\text{If indep. } Cov(X, Y) = 0, \rho_{XY} = 0$$
$$Var(aX + bY) =$$
$$a^2 VarX + b^2 VarY + 2abCov(X, Y)$$

Some integrals

$$\Gamma(\alpha) = \int_0^{\infty} t^{\alpha-1} e^{-t} dt$$
$$\int_0^{\infty} e^{-z^2/2} dz = \sqrt{\frac{\pi}{2}}$$
$$B(\alpha, \beta) = \int_0^1 x^{\alpha-1}(1 - x)^{\beta-1}dx$$
$$\int e^{ax}dx = (1/a)e^{ax} + C$$
$$\int udv = uv - \int vdu$$
$$\int x^n dx = (x^{n+1})/(n + 1)$$

Additional Identities

$$\sum_{x=1}^{\infty} a^{x-1} = \frac{1}{1 - a}$$
$$\Gamma(\alpha + 1) = \alpha\Gamma(\alpha)$$
$$\Gamma(n) = (n - 1)!$$
$$B(\alpha, \beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha + \beta)}$$
$$\lim_x \to 0^+ \log_b(x) = -\infty$$
$$\log(xy) = \log(x) + \log(y)$$
$$\log(x/y) = \log(x) - \log(y)$$
$$\log(x^y) = y \log x$$
$$a^n/a^m = a^{n-m}$$
$$a^n a^m = a^{n+m}$$

Notes

$COV(X, Y) = 0 \not\Rightarrow$ Independence

$E[X] = 0$ if f(x) symmetric about 0

$P(X < Y) \implies$ line at X=Y w up. reg. shaded

$P(X > Y) \implies$ line at X=Y w low. reg.shaded

Even function: $f(-x) = f(x)$

Not indep. if the support of (X,Y) is NOT a rect. but being rect. DNI independence

Remember that

$p(x, y) = p(X = x \cap Y = y)$

Random Sample and Sums of Random Variables

A random sample is iid

$$f(x_1, \cdots, x_n|\theta) = \prod_{i=1}^{n} f(x_i|\theta) \text{ mutual indep.}$$

Rand. samp. from a finite pop. w/o rep. not indep.

if N is large compared to n then nearly indep.

$$P(X_1 > 2, \cdots, P(X_n > 2)) = [P(X_1 > 2)]^n$$

$$\min_a \sum_{i=1}^{n}(x_i - a)^2 = \sum_{i=1}^{n}(x_i - \overline{x})^2$$

$$(n-1)^2 s^2 = \sum_{i=1}^{n}(x_i - \overline{x})^2 = \sum_{i=1}^{n} x_i^2 - n\overline{x}^2$$

$$\mathbb{E}\left(\sum_{i=1}^{n} g(X_i)\right) = n\mathbb{E}\left(g(X_1)\right)$$

$$\text{Var}\left(\sum_{i=1}^{n} g(X_i)\right) = n\text{Var}\left(g(X_1)\right)$$

$$\mathbb{E}(\overline{X}) = \mu$$

$$\text{Var}\overline{X} = \sigma^2/n$$

$$ES^2 = \sigma^2$$

$$M_{\overline{X}}(t) = [M_X(t/n)]^2$$

$$Z_i \sim \text{Cauchy}(0,1) \implies \sum Z_i \sim \text{Cauchy(0, n)}$$

$$\overline{Z} \sim \text{Cauchy}(0,1)$$

Sampling from loc-scale families: $\overline{X} = \sigma\overline{Z} + \mu$

Sampling from the Normal distribution

Let $X_1, \cdots, X_n$ be a rand. samp. from $n(\mu, \sigma^2)$

$\overline{X}$ and S^2 are indep.

$$\overline{X} \sim n(\mu, \sigma^2/n)$$

$$(n-1)S^2/\sigma^2 \sim \chi^2_{n-1}$$

$$Z \sim n(0,1) \implies Z^2 \sim \chi^2_1$$

$$X_i \sim \chi^2_{p_i} \implies \sum X_i \sim \chi^2_{p_1, \cdots, p_n}$$

$$\overline{X} - \mu/(\sigma/\sqrt{n}) \sim n(0,1)$$

$$\overline{X} - \mu/(\sigma/\sqrt{n})/(\sqrt{S^2/\sigma^2})$$

numerator is n(0,1) and denom. is $\sqrt{\chi^2_{n-1}/n-1}$

$$\overline{X} - \mu/(S/\sqrt{n}) \sim t_{n-1}$$

$$(S^2_X/\sigma^2_X)/(S^2_Y/\sigma^2_Y) \sim F_{n-1,m-1}$$

$$X\ simF_{p,q} \implies 1/X \sim F_{q,p}$$

$$X \sim t_q \implies X^2 \sim F_{1,q}$$

$$X \sim F_{p,q} \implies (p/q)X/(1+(p/q)X) \sim beta(p/2, q/2)$$

Order statistics

If (iid)$X_i \sim F(x|\theta) \implies P(X_{(n) \leq x}) = [F(x|\theta)]^n$

and $P(X_{(1)} \leq x) = 1 - [1 - F(x|\theta)]^n$

$F_X(x)$ is right-continuous: $\lim_{x \downarrow x_0} F(x) = F(X_0)$

Outside of the exp. fam. it is rare to have a suff. stat. smaller than the dim. of the size of the sample

Baye's Estimation

$$f(\theta|x) = f(x|\theta)f(\theta)/f(x) \implies f(\theta|x) \propto f(x|\theta)f(\theta)$$

Delta Method

Y_n a seq. satisfies $\sqrt{n}(Y_n - \theta) \to n(0, \sigma^2)$

$$\sqrt{n}\left[g(Y_n) - g(\theta)\right] \to n(0, \sigma^2[g^{'}(\theta)]^2)$$

Convergence Concepts

Convergence in probability:

$Z_1, \cdots$(seq. of r.v.s) as n grows large, $Z_n \to c$

$$Z_n \in (c - \epsilon, c + \epsilon)$$

$P(|Z_n - c| < \epsilon)$ as n grows, $P(|Z_n - c| < \epsilon) = 1$

n grows, need:$\mathbb{E}(Zn) \to c$ and $Var(Z_n) \to 0$

Convergence in distribution:

If $Z_n \sim F_n(z)$we have cov. in dist. if $F_n(z) \to F(z) \forall z \in \mathbb{R}$ F is cont. then $Z_n \to Z$ in dist.

$\sqrt{n}(\overline{X}_n - \mu)/\sigma$ limiting st. norm. (CLT)

$$\mathbb{E}(Z_n) \approx \mathbb{E}(Z)$$

$$\text{Var}(Z_n) \approx \text{Var}(Z)$$

$$P(a < Z_n < b) \approx P(a < Z < b)$$

WLLN: conv. in p. $\lim_{n\to\infty} P(|\overline{Z}_n - \mu| < \epsilon) = 1$

SLLN: alm. surely$P(\lim_{n\to\infty} |\overline{Z}_n - \mu| < \epsilon) = 1$

Slutsky's THM: if $X_n \to X$ in dist.

and $Y_n \to a \implies Y_n X_n \to aX$ in dist.

and $\implies Y_n + X_n \to a + X$ in dist.

Sufficient Statistics

T(X) is a suff. stat. for θ if the conditional distribution of the sample does not dep. on θ

Factorization THM: $f(\vec{x}|\theta) = g(T(\vec{x}|\theta)h(\vec{x})$

$$f(x|\theta) = h(x)c(\theta)\exp\left(\sum_{i-1}^{k} w_i(\theta)t_i(x)\right)$$

$$\implies T(X) = \left(\sum_{j=1}^{n} t_1(X_j), \cdots, \sum_{j=1}^{n} t_k(X_j)\right)$$

$$f(x|\theta) = L(\theta|\vec{x}) = \prod_{i=1}^{n} f(x_i|\theta)$$

An unbiased estimator:$E(\hat{\theta}) = \theta$

M.o.M: set $(1/n)\sum_{i=1}^{n} x_i = E[X] = \mu$

$$(1/n)\sum_{i=1}^{n} x_i^2 = E[X^2]] = \sigma^2 + \mu^2$$

$$\text{Bias} = E[\tilde{\theta}] - \theta$$

$$MSE(\tilde{\theta}) = var(\tilde{\theta}) + Bias^2(\tilde{\theta})$$

If MSE($\hat{\theta}$) < MSE($\tilde{\theta}$) prefer $\hat{\theta}$

MLEs may be unbiased but they will have min. var. then you turn them into MVUE

Hypothesis Testing

P(Type I Error) = P(Rej H0|H0 is T) = α

P(Type II Error) = P(not rej H0|H0 is F) = β

as alp goes down beta goes up

as n goes up beta goes down

power = $1 - \beta = P$(rej H0|H0 is F)

Piv. quant. : a test stat. that doesnt dep. on theta

H0: mu = mu0 vs. H1: mu > mu0

$$T = (\overline{X} - \mu_0)/(\sigma/\sqrt{n}) \sim n(0,1)$$

Rej. if T $> Z_\alpha$if lower tail $T < -Z_\alpha$

If two tailed: rej when $|T| > z_{\alpha/2}$

When sig-sqrd is unk:$T = (\overline{X} - \mu_0)/(S/\sqrt{n}) \sim t_{n-1}$

If n $> 30 \implies t_{n-1} \to n(0,1)$

For H0: $\sigma^2 = \sigma^2_0$ vs. H1: $\sigma^2 > \sigma^2_0$

$$T = (n-1)S^2/\sigma^2_0 \sim \chi^2_{n-1}$$

If any other test, use CLT and standard error.

LRT and UMP

Neyman-Pearson Lemma: use for simple hypotheses.

$$H0 : \theta = \theta_0 \text{ vs } H1 : \theta = \theta_1$$

If $\exists k > 0$ and a region c s.t.

$$P[(x_1, \cdots, x_n) \in RR|\theta = \theta_0] = \alpha$$

$$L(\theta_0)/L(\theta_1) \leq k \text{ for } (x_1, \cdots, x_n) \in RR$$

$$L(\theta_0)/L(\theta_1) \geq k \text{ for } (x_1, \cdots, x_n) \notin RR$$

Then c is a best (UMP) RR of size alp for the test

Notes

When samples are equally likley to be drawn:

Prob. that some set becomes a samp.: $1/\binom{N}{n}$

When samples are drawn iid:

Prob. that some set j becomes a samp.: n/N

Prob. that some set i and j becomes a samp.:

$$(n/N)/(n-1/N-1)$$

$$X, Y \sim Poisson(\lambda) \implies X + Y \sim Poisson(2\lambda)$$

$$\lim_{n\to\infty} P(|X - \mu| > \epsilon) \leq \lim_{n\to\infty} Var(\overline{X}/\epsilon^2)$$

Standard deviation or st. error:

$\overline{X}$	$\sigma/\sqrt{n}$	$s/\sqrt{n}$
p	$\sqrt{P(1-P)/n}$	$\sqrt{p(1-p)/n}$
$\overline{X}_1 - \overline{X}_1$	$\sqrt{\sigma^2_1/n_1 + \sigma^2_2/n_2}$	$\sqrt{s^2_1/n_1 + s^2_2/n_2}$
$p_1 - p_2$	$\sqrt{P_1(1-P_1)/n_1 - P_2(1-P)/n_2}$	