STAT 512 Summary Sheet

Karl B. Gregory

1. Transformations of a random variable

   - Let X be a rv with support $\mathcal{X}$ and let g be a function mapping $\mathcal{X}$ to $\mathcal{Y}$ with inverse mapping

   $$g^{-1}(A) = \{x \in \mathcal{X} : g(x) \in A\} \quad \text{for all} \quad A \subset \mathcal{Y}.$$

   - If X is a discrete rv, the rv $Y = g(X)$ is discrete and has pmf given by

   $$p_Y(y) = \begin{cases} \sum_{x \in g^{-1}(y)} p_X(x) & \text{for } y \in \mathcal{Y} \\ 0 & \text{for } y \notin \mathcal{Y}. \end{cases}$$

   - If X has pdf f_X then the cdf of $Y = g(X)$ is given by

   $$F_Y(y) = \int_{\{g(x) < y\}} f_X(x)dx, \quad -\infty < y < \infty.$$

   - If X has pdf f_X which is continuous on $\mathcal{X}$ and if g is monotone and g^{-1} has a continuous derivative then the pdf of $Y = g(X)$ is given by

   $$f_Y(y) = \begin{cases} f_X(g^{-1}(y))|\frac{d}{dy}g^{-1}(y)| & \text{for } y \in \mathcal{Y} \\ 0 & \text{for } y \notin \mathcal{Y}. \end{cases}$$

   - If X has mgf $M_X(t)$, the mgf of $Y = aX + b$ is given by $M_Y(t) = e^{bt}M_X(at)$.
   - If X has continuous cdf F_X, the rv $U = F_X(X)$ has the uniform distribution on $(0, 1)$.
   - To generate $X \sim F_X$, generate $U \sim \text{Uniform}(0, 1)$ and set $X = F_X^{-1}(U)$, where

   $$F_X^{-1}(u) = \inf\{x : F_X(x) \geq u\}.$$

   If F_X is strictly increasing, F_X^{-1} satisfies

   $$x = F_X^{-1}(u) \iff u = F_X(x)$$

   for all $0 < u < 1$ and $-\infty < x < \infty$.

2. Transformations of multiple random variables

- Let (X_1, X_2) be a pair of continuous random variables with pdf f_{X_1,X_2} having joint support
$$\mathcal{X} = \{(x_1, x_2) : f_{X_1,X_2}(x_1, x_2) > 0\}$$
and let $Y_1 = g_1(X_1, X_2)$ and $Y_2 = g_2(X_1, X_2)$, where g_1 and g_2 form a one-to-one transformation of $\mathcal{X}$ onto the set
$$\mathcal{Y} = \{(y_1, y_2) : y_1 = g_1(x_1, x_2), y_2 = g_2(x_1, x_2), (x_1, x_2) \in \mathcal{X}\}.$$
Then the joint pdf f_{Y_1,Y_2} of rv pair (Y_1, Y_2) is given by
$$f_{Y_1,Y_2}(y_1, y_2) = \begin{cases} f_{X_1,X_2}(g_1^{-1}(y_1, y_2), g_2^{-1}(y_1, y_2))|J(y_1, y_2)| & \text{for } (y_1, y_2) \in \mathcal{Y} \\ 0 & \text{for } (y_1, y_2) \notin \mathcal{Y} \end{cases},$$
where g_1^{-1} and g_2^{-1} are the functions satisfying
$$\begin{array}{c} y_1 = g_1(x_1, x_2) \\ y_2 = g_2(x_1, x_2) \end{array} \iff \begin{array}{c} x_1 = g_1^{-1}(y_1, y_2) \\ x_2 = g_2^{-1}(y_1, y_2) \end{array}$$
and
$$J(y_1, y_2) = \begin{vmatrix} \frac{\partial}{\partial y_1} g_1^{-1}(y_1, y_2) & \frac{\partial}{\partial y_2} g_1^{-1}(y_1, y_2) \\ \frac{\partial}{\partial y_1} g_2^{-1}(y_1, y_2) & \frac{\partial}{\partial y_2} g_2^{-1}(y_1, y_2) \end{vmatrix},$$
provided $J(y_1, y_2)$ is not equal to zero for all $(y_1, y_2) \in \mathcal{Y}$.

- For real numbers a, b, c, d,
$$\begin{vmatrix} a & b \\ c & d \end{vmatrix} = ad - bc.$$

- Let $X_1, \ldots, X_n$ be mutually independent rvs such that X_i has mgf $M_{X_i}(t)$ for $i = 1, \ldots, n$. Then the mgf of $Y = X_1 + \cdots + X_n$ is given by
$$M_Y(t) = \prod_{i=1}^{n} M_{X_i}(t).$$
It follows that if $X_1, \ldots, X_n$ are independent identically distributed rvs, each with mgf $M_X(t)$, then the mgf of $Y = X_1 + \cdots + X_n$ is given by
$$M_Y(t) = [M_X(t)]^n.$$

3. Random samples, sums of rvs, and order statistics

- A *random sample* is a collection of mutually independent rvs all having the same distribution. This common distribution is called the *population distribution* and quantities describing the population distribution are called *population parameters*. A *sample statistic* is a function of the rvs in the random sample and its distribution is called its *sampling distribution*. What we can learn about population parameters from sample statistics has everything to do with their sampling distributions.

- Random samples might be introduced in any of the following ways,
  - Let $X_1, \ldots, X_n$ be a random sample from a population with such-and-such distribution.
  - Let $X_1, \ldots, X_n$ be independent copies of the random variable X, where X has such-and-such distribution.
  - Let $X_1, \ldots, X_n$ be iid rvs from such-and-such distribution (iid means *independent and identically distributed*).
  - Let $X_1, \ldots, X_n \overset{\text{ind}}{\sim}$ a distribution.
- Let $X_1, \ldots, X_n$ be a random sample from a population with mean μ and variance $\sigma^2 < \infty$ and define the sample mean and sample variance as

$$\bar{X}_n = \frac{1}{n}(X_1 + \cdots + X_n)$$
$$S_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2.$$

  Then we have $\mathbb{E}\bar{X}_n = \mu$, $\operatorname{Var}\bar{X}_n = \sigma^2/n$, and $\mathbb{E}S_n^2 = \sigma^2$.
- The *order statistics* of a rs $X_1, \ldots, X_n$ are the ordered data values

$$X_{(1)} < X_{(2)} < \cdots < X_{(n)}.$$

- Let $X_{(1)} < \cdots < X_{(n)}$ be the order statistics of a random sample from a population with cdf F_X and pdf f_X. Then the pdf of the kth order statistic $X_{(k)}$ is given by

$$f_{X_{(k)}}(x) = \frac{n!}{(k-1)!(n-k)!}[F_X(x)]^{k-1}[1 - F_X(x)]^{n-k} f_X(x)$$

  for $k = 1, \ldots, n$. In particular, the minimum $X_{(1)}$ and the maximum $X_{(n)}$ have the pdfs

$$f_{X_{(1)}}(x) = n[1 - F_X(x)]^{n-1} f_X(x)$$
$$f_{X_{(n)}}(x) = n[F_X(x)]^{n-1} f_X(x)$$

- Let $X_{(1)} < \cdots < X_{(n)}$ be the order statistics of a random sample from a population with cdf F_X and pdf f_X. Then the joint pdf of the jth and kth order statistics $(X_{(j)}, X_{(k)})$, with $1 \le j < k \le n$, is given by

$$f_{X_{(j)},X_{(k)}}(u,v) = \frac{n!}{(j-1)!(k-1-j)!(n-k)!} f_X(u) f_X(v)[F_X(u)]^{j-1}[F_X(v) - F_X(u)]^{k-1-j}[1 - F_X(v)]^{n-k}$$

  for $-\infty < u < v < \infty$.

  In particular, the joint pdf of the minimum and maximum $(X_{(1)}, X_{(n)})$ is

$$f_{X_{(1)},X_{(n)}}(u,v) = n(n-1) f_X(u) f_X(v)[F_X(v) - F_X(u)]^{n-2}$$

  for $-\infty < u < v < \infty$.

4. Pivot quantities and sampling from the Normal distribution

- A *pivot quantity* is a function of sample statistics and population parameters which has a known distribution (not depending on unknown parameters).
- A $(1-\alpha)100\%$ *confidence interval* (CI) for an unknown parameter $\theta \in \mathbb{R}$ is an interval (L, U) where L and U are random variables such that $P(L < \theta < U) = 1 - \alpha$, for $\alpha \in (0, 1)$.
- Important distributions and upper quantile notation:
  - The $\text{Normal}(0, 1)$ distribution has pdf

$$\phi(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2} \mathbb{1}(-\infty < z < \infty),$$

    and for $\xi \in (0, 1)$ we denote by z_ξ the value satisfying $P(Z > z_\xi) = \xi$, where $Z \sim \text{Normal}(0, 1)$.
  - The Chi-square distributions χ^2_ν, $\nu = 1, 2, \dots$ have the pdfs

$$f_X(x; \nu) = \frac{1}{\Gamma(\nu/2) 2^{\nu/2}} x^{\nu/2 - 1} e^{-x/2} \mathbb{1}(x > 0),$$

    and for $\xi \in (0, 1)$ we denote by $\chi^2_{\nu,\xi}$ the value satisfying $P(X > \chi^2_{\nu,\xi}) = \xi$, where $X \sim \chi^2_\nu$. The parameter ν is called the *degrees of freedom*.
  - The t distributions t_ν, $\nu = 1, 2, \dots$ have the pdfs

$$f_T(t; \nu) = \frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{\nu}{2})} \frac{1}{\sqrt{\nu\pi}} \left(1 + \frac{t^2}{\nu}\right)^{-(\nu+1)/2} \mathbb{1}(-\infty < t < \infty),$$

    and for $\xi \in (0, 1)$ we denote by $t_{\nu,\xi}$ the value satisfying $P(T > t_{\nu,\xi}) = \xi$, where $T \sim t_\nu$. The parameter ν is called the *degrees of freedom*.
  - The F distributions F_{ν_1,ν_2}, $\nu_1, \nu_2 = 1, 2, \dots$ have the pdfs

$$f_R(r; \nu_1, \nu_2) = \frac{\Gamma(\frac{\nu_1+\nu_2}{2})}{\Gamma(\frac{\nu_1}{2})\Gamma(\frac{\nu_2}{2})} \left(\frac{\nu_1}{\nu_2}\right)^{\nu_1/2} r^{(\nu_1-2)/2} \left(1 + \frac{\nu_1}{\nu_2} r\right)^{-(\nu_1+\nu_2)/2} \mathbb{1}(r > 0),$$

    and for $\xi \in (0, 1)$ we denote by $F_{\nu_1,\nu_2,\xi}$ the value satisfying $P(R > F_{\nu_1,\nu_2,\xi}) = \xi$, where $R \sim F_{\nu_1,\nu_2}$. The parameters ν_1 and ν_2 are called, respectively, the *numerator degrees of freedom* and the *denominator degrees of freedom*.
- Let $Z \sim \text{Normal}(0, 1)$ and $X \sim \chi^2_\nu$ be independent rvs. Then

$$T = \frac{Z}{\sqrt{X/\nu}} \sim t_\nu.$$

- Let $X_1 \sim \chi^2_{\nu_1}$ and $X_2 \sim \chi^2_{\nu_2}$ be independent rvs. Then

$$R = \frac{X_1/\nu_1}{X_2/\nu_2} \sim F_{\nu_1,\nu_2}.$$

- Let $X_1, \dots, X_n \overset{\text{ind}}{\sim} \text{Normal}(\mu, \sigma^2)$ distribution. Then
  - $\sqrt{n}(\bar{X}_n - \mu)/\sigma \sim \text{Normal}(0,1)$, giving $P\left(\bar{X}_n - z_{\alpha/2}\dfrac{\sigma}{\sqrt{n}} < \mu < \bar{X}_n + z_{\alpha/2}\dfrac{\sigma}{\sqrt{n}}\right) = 1-\alpha.$
  - $(n-1)S_n^2/\sigma^2 \sim \chi^2_{n-1}$, giving $P\left(\dfrac{(n-1)S_n^2}{\chi^2_{n-1,\alpha/2}} < \sigma^2 < \dfrac{(n-1)S_n^2}{\chi^2_{n-1,1-\alpha/2}}\right) = 1-\alpha.$
  - $\sqrt{n}(\bar{X}_n - \mu)/S_n \sim t_{n-1}$, giving $P\left(\bar{X}_n - t_{n-1,\alpha/2}\dfrac{S_n}{\sqrt{n}} < \mu < \bar{X}_n + t_{n-1,\alpha/2}\dfrac{S_n}{\sqrt{n}}\right) = 1-\alpha.$
- Consider two independent random samples $X_1, \dots, X_{n_1} \overset{\text{ind}}{\sim} \text{Normal}(\mu_1, \sigma_1^2)$ and $Y_1, \dots, Y_{n_2} \overset{\text{ind}}{\sim} \text{Normal}(\mu_2, \sigma_2^2)$ with sample means $\bar{X}_{n_1}$ and $\bar{Y}_{n_2}$ and sample variances S_1^2 and S_2^2, respectively. Then
$$\frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} \sim F_{n_1-1,n_2-1},$$
giving
$$P\left(\frac{S_2^2}{S_1^2}F_{n_1-1,n_2-1,1-\alpha/2} < \frac{\sigma_2^2}{\sigma_1^2} < \frac{S_2^2}{S_1^2}F_{n_1-1,n_2-1,\alpha/2}\right) = 1-\alpha.$$

5. Parametric estimation and properties of estimators

- *Parametric framework*: Let $X_1, \dots, X_n$ have a joint distribution which depends on a finite number of parameters $\theta_1, \dots, \theta_d$, the values of which are unknown and lie in the spaces $\Theta_1, \dots, \Theta_d$, respectively, where $\Theta_k \subset \mathbb{R}$, for $k = 1, \dots, d$. To know the joint distribution of $X_1, \dots, X_n$, it is sufficient to know the values of $\theta_1, \dots, \theta_d$, so we define estimators $\hat{\theta}_1, \dots, \hat{\theta}_d$ of $\theta_1, \dots, \theta_d$ based on $X_1, \dots, X_n$.
- *Nonparametric framework*: Let $X_1, \dots, X_n$ have a joint distribution which depends on an infinite number of parameters. For example, suppose $X_1, \dots, X_n$ is a random sample from a distribution with pdf f_X, where we do not specify any functional form for f_X. Then we may regard the value of the pdf $f_X(x)$ at every point $x \in \mathbb{R}$ as a parameter, and since there are an infinite number of values of $x \in \mathbb{R}$, there are infinitely many parameters.
- The *bias* of an estimator $\hat{\theta}$ of a parameter θ is defined as
$$\text{Bias}\,\hat{\theta} = \mathbb{E}\hat{\theta} - \theta.$$
And estimator is called *unbiased* if $\text{Bias}\,\hat{\theta} = 0$, that is if $\mathbb{E}\hat{\theta} = \theta$.
- The *standard error* of an estimator $\hat{\theta}$ of a parameter θ is defined as
$$\text{SE}\,\hat{\theta} = \sqrt{\text{Var}\,\hat{\theta}},$$
so the standard error of an estimator is simply its standard deviation.
- The *mean squared error* (MSE) of an estimator $\hat{\theta}$ of a parameter θ is defined as
$$\text{MSE}\,\hat{\theta} = \mathbb{E}(\hat{\theta} - \theta)^2,$$
so the MSE of $\hat{\theta}$ is the expected squared distance between $\hat{\theta}$ and θ.

- $\text{MSE}\,\hat\theta = \text{Var}\,\hat\theta + (\text{Bias}\,\hat\theta)^2$.
- If $\hat\theta$ is an unbiased estimator of θ, it is *not* generally true that $\tau(\hat\theta)$ will be an unbiased estimator of $\tau(\theta)$.

6. Large-sample properties of estimators, consistency and WLLN

- A sequence of estimators $\{\hat\theta_n\}_{n\geq 1}$ of a parameter $\theta \in \Theta \subset \mathbb{R}$ is called *consistent* if for every $\epsilon > 0$ and every $\theta \in \Theta$,
$$\lim_{n\to\infty} P(|\hat\theta_n - \theta| < \epsilon) = 1.$$
Note: When we write $\hat\theta_n$, with n in the subscript, it is understood that we are considering a sequence of estimators for all sample sizes $n \geq 1$.
- We often express that $\hat\theta$ is a consistent estimator of θ by writing
$$\hat\theta_n \xrightarrow{\text{p}} \theta,$$
where $\xrightarrow{\text{p}}$ denotes something called *convergence in probability*. Saying "$\hat\theta_n$ converges in probability to θ" is equivalent to saying "$\hat\theta$ is a consistent estimator of θ".
- The *weak law of large numbers* (WLLN): Let $X_1, \dots, X_n$ be a rs from a distribution with mean μ and variance $\sigma^2 < \infty$. Then $\bar X_n = n^{-1}(X_1 + \cdots + X_n)$ is a consistent estimator of μ.
- The estimator $\hat\theta_n$ is a consistent estimator for θ if
  (a) $\lim_{n\to\infty} \text{Var}\,\hat\theta_n = 0$
  (b) $\lim_{n\to\infty} \text{Bias}\,\hat\theta_n = 0$

  So we can show that $\hat\theta_n$ is a consistent estimator of θ by showing $\text{MSE}\,\hat\theta_n \to 0$ as $n \to \infty$.
- Let $\{\hat\theta_{1,n}\}_{n\geq 1}$ and $\{\hat\theta_{2,n}\}_{n\geq 1}$ be sequences of estimators for θ_1 and θ_2, respectively. Then
  (a) $\hat\theta_{1,n} \pm \hat\theta_{2,n}$ is consistent for $\theta_1 \pm \theta_2$.
  (b) $\hat\theta_{1,n} \cdot \hat\theta_{2,n}$ is consistent for $\theta_1 \cdot \theta_2$.
  (c) $\hat\theta_{1,n}/\hat\theta_{2,n}$ is consistent for θ_1/θ_2, so long as $\theta_2 \neq 0$.
  (d) For any continuous function $\tau : \mathbb{R} \to \mathbb{R}$, $\tau(\hat\theta_{1,n})$ is consistent for $\tau(\theta_1)$.
  (e) For any sequences $\{a_n\}_{n\geq 1}$ and $\{b_n\}_{n\geq 1}$ such that $\lim_{n\to\infty} a_n = 1$ and $\lim_{n\to\infty} b_n = 0$, $a_n\hat\theta_{1,n} + b_n$ is consistent for θ_1.
- If $X_1, \dots, X_n$ is a random sample from a distribution with mean μ and variance $\sigma^2 < \infty$ and 4th moment $\mu_4 < \infty$, the sample variance S_n^2 is consistent for σ^2.
- If $X_1, \dots, X_n$ is a random sample from the $\text{Bernoulli}(p)$ distribution and $\hat p = n^{-1}(X_1 + \cdots + X_n)$, then $\hat p(1 - \hat p)$ is a consistent estimator of $p(1-p)$.

7. Large-sample pivot quantities and central limit theorem

- A sequence of random variables $Y_1, Y_2, \dots$ with cdfs $F_{Y_1}, F_{Y_2}, \dots$ is said to *converge in distribution* to the random variable $Y \sim F_Y$ if
$$\lim_{n\to\infty} F_{Y_n}(y) = F_Y(y)$$

for all $y \in \mathbb{R}$ at which F_Y is continuous. We express convergence in distribution with the notation $Y_n \xrightarrow{\text{D}} Y$ and we refer to the distribution with cdf F_Y as the *asymptotic distribution* of the sequence of random variables $Y_1, Y_2, \dots$

- *Convergence in distribution* is a sense in which a random variable (we can think of a quantity computed on larger and larger samples) behaves more and more like another random variable (as the sample size grows).
- A *large-sample pivot quantity* is a function of sample statistics and population parameters for which the asymptotic distribution is fully known (not depending on unknown parameters).
- The *Central Limit Theorem* (CLT): Let $X_1, \dots, X_n$ be a random sample from *any* distribution and suppose the distribution has mean μ and variance $\sigma^2 < \infty$. Then

$$\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \xrightarrow{\text{D}} Z, \quad Z \sim \text{Normal}(0,1).$$

Application to CIs: The above means that for any $\alpha \in (0,1)$,

$$\lim_{n\to\infty} P\left(-z_{\alpha/2} < \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} < z_{\alpha/2}\right) = 1 - \alpha,$$

giving that

$$\bar{X}_n \pm z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

is an approximate $(1-\alpha)100\%$ CI for μ as long as n is large. A rule of thumb says $n \geq 30$ is "large".

- *Slutzky's Theorem*: If $X_n \xrightarrow{\text{D}} X$ and $Y_n \xrightarrow{\text{p}} a$, then

$$\begin{aligned} X_n + Y_n &\xrightarrow{\text{D}} X + a. \\ X_n Y_n &\xrightarrow{\text{D}} Xa \\ X_n / Y_n &\xrightarrow{\text{D}} X/a, \quad \text{provided } a \neq 0. \end{aligned}$$

- *Corollary of Slutzky's Theorem*: Let $X_1, \dots, X_n$ be a rs from *any* distribution and suppose the distribution has mean μ and variance $\sigma^2 < \infty$. Moreover, let $\hat{\sigma}_n$ be a consistent estimator of σ. Then

$$\frac{\bar{X}_n - \mu}{\hat{\sigma}_n/\sqrt{n}} \xrightarrow{\text{D}} Z, \quad Z \sim \text{Normal}(0,1).$$

Application to CIs: The above means that for any $\alpha \in (0,1)$,

$$\lim_{n\to\infty} P\left(-z_{\alpha/2} < \frac{\bar{X} - \mu}{\hat{\sigma}_n/\sqrt{n}} < z_{\alpha/2}\right) = 1 - \alpha,$$

giving that

$$\bar{X}_n \pm z_{\alpha/2}\frac{\hat{\sigma}_n}{\sqrt{n}}$$

is an approximate $(1-\alpha)100\%$ CI for μ as long as n is large.

A rather arbitrary rule of thumb says $n \geq 30$ is "large".

- Let $X_1, \ldots, X_n$ be a random sample from *any* distribution and suppose the distribution has mean μ and variance $\sigma^2 < \infty$ and 4th moment $\mu_4 < \infty$. Then

$$\frac{\bar{X}_n - \mu}{S_n/\sqrt{n}} \xrightarrow{\text{D}} Z, \quad Z \sim \text{Normal}(0,1),$$

so that for large n,

$$\bar{X}_n \pm z_{\alpha/2}\frac{S_n}{\sqrt{n}}$$

is an approximate $(1-\alpha)100\%$ CI for μ.

- Let $X_1, \ldots, X_n$ be a random sample from the $\text{Bernoulli}(p)$ distribution and let $\hat{p}_n = \bar{X}_n$. Then

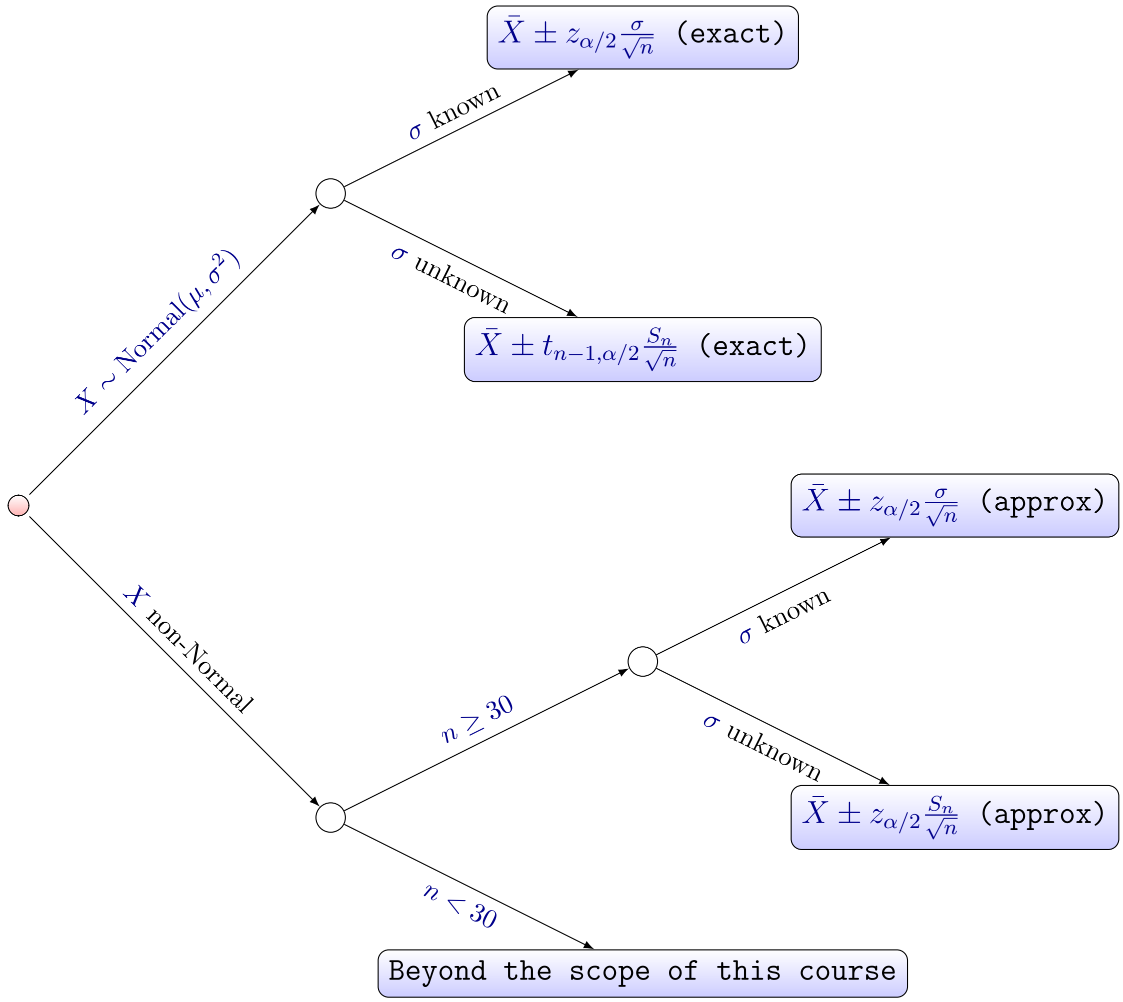
$$\frac{\hat{p}_n - p}{\sqrt{\frac{\hat{p}_n(1-\hat{p}_n)}{n}}} \xrightarrow{\text{D}} Z, \quad Z \sim \text{Normal}(0,1),$$

so that for large n (a rule of thumb is to require $\min\{n\hat{p}_n, n(1-\hat{p}_n)\} \geq 15$),

$$\hat{p}_n \pm z_{\alpha/2}\sqrt{\frac{\hat{p}_n(1-\hat{p}_n)}{n}}$$

is an approximate $(1-\alpha)100\%$ CI for p.

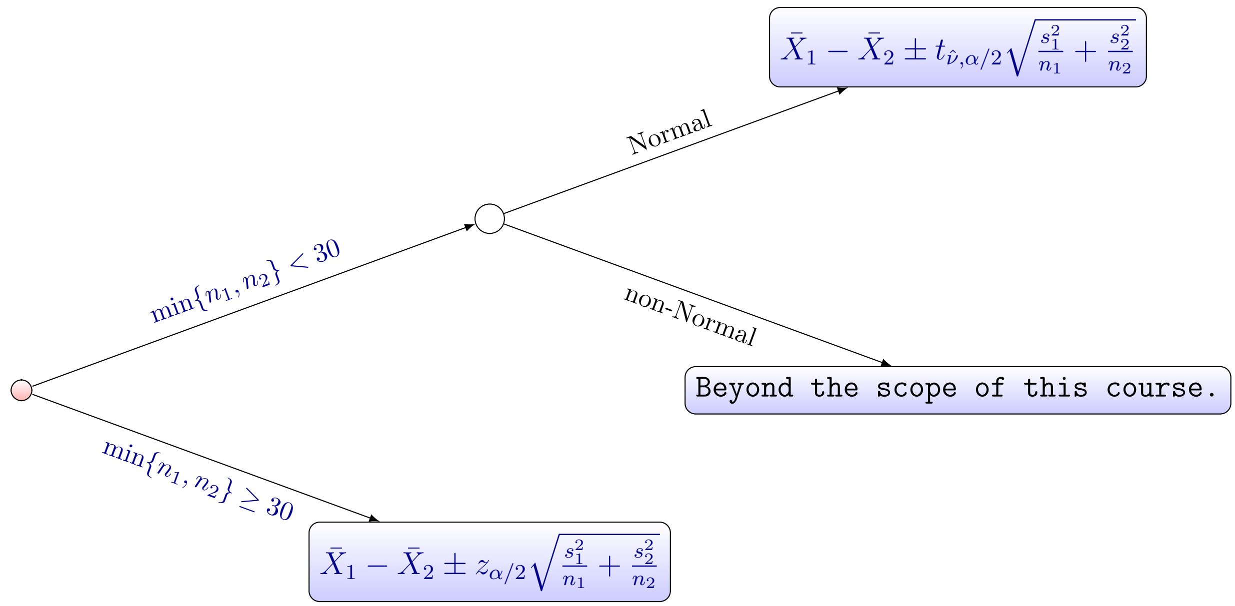
- Summary of $(1-\alpha)100\%$ CIs for μ: Let $X_1, \ldots, X_n$ be independent rvs with the same distribution as the rv X, where $\mathbb{E}X = \mu$, $\text{Var}\,X = \sigma^2$.



8. Classical two-sample results comparing means, proportions, and variances

   - Summary of $(1-\alpha)100\%$ CIs for $\mu_1 - \mu_2$: Let $X_1, \ldots, X_{n_1}$ be independent rvs with the same distribution as the rv X, where $\mathbb{E}X = \mu_1$ and $\operatorname{Var} X = \sigma_1^2$ and let $Y_1, \ldots, Y_{n_2}$ be independent rvs with the same distribution as the rv Y, where $\mathbb{E}Y = \mu_2$ and $\operatorname{Var} \sigma_2^2$.

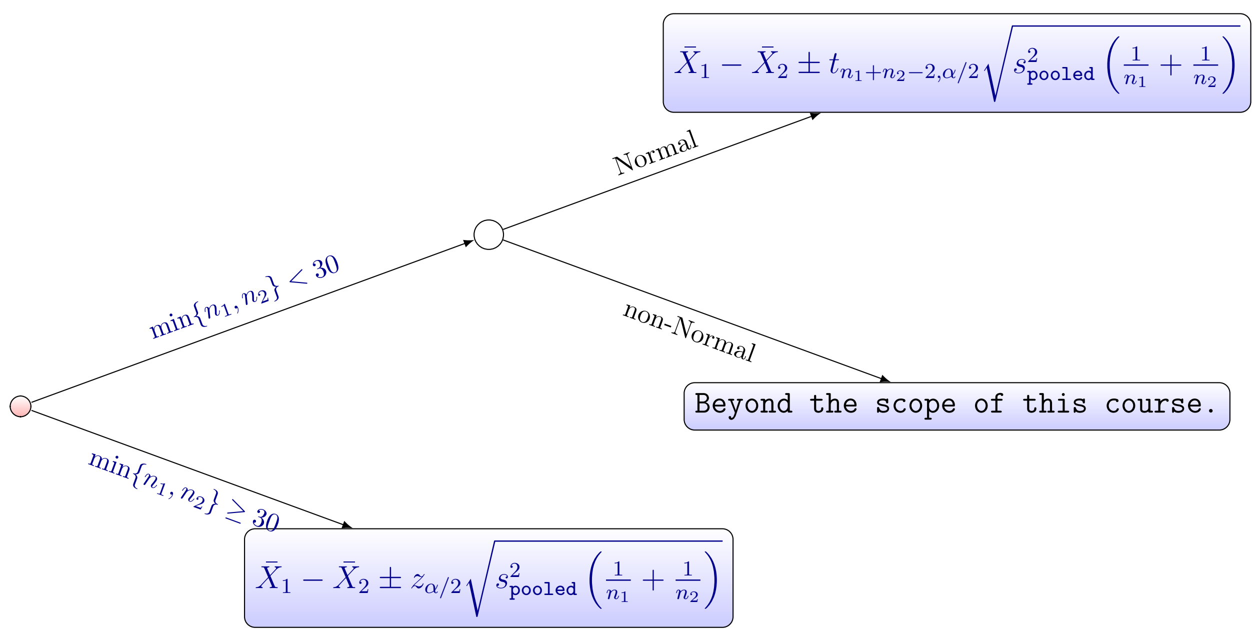
Case (i): $\sigma_1^2 \neq \sigma_2^2$:



In the above

$$\hat{\nu} = \left(\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}\right)^2 \left[\frac{\left(\frac{S_1}{n_1}\right)^2}{n_1 - 1} + \frac{\left(\frac{S_2^2}{n_2}\right)^2}{n_2 - 1}\right]^{-1}.$$

Case (ii): $\sigma_1^2 = \sigma_2^2$:



- A $(1-\alpha)100\%$ CI for $p_1 - p_2$: Let $X_1, \ldots, X_{n_1} \stackrel{\text{ind}}{\sim} \text{Bernoulli}(p_1)$ and $Y_1, \ldots, Y_{n_2} \stackrel{\text{ind}}{\sim} \text{Bernoulli}(p_2)$ and let $\hat{p}_1 = n_1^{-1}(X_1 + \cdots + X_{n_1})$ and $\hat{p}_2 = n_2^{-1}(Y_1 + \cdots + Y_{n_2})$. Then for large n_1 and n_2 (Rule of thumb is $\min\{n_1\hat{p}_1, n_1(1-\hat{p}_1)\} \geq 15$ and $\min\{n_2\hat{p}_2, n_2(1-\hat{p}_2)\} \geq 15$)

$$\hat{p}_1 - \hat{p}_2 \pm z_{\alpha/2}\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$$

  is an approximate $(1-\alpha)100\%$ CI for $p_1 - p_2$.

- A $(1-\alpha)100\%$ CI for σ_2^2/σ_1^2: Consider two independent random samples $X_1, \ldots, X_{n_1} \stackrel{\text{ind}}{\sim} \text{Normal}(\mu_1, \sigma_1^2)$ and $Y_1, \ldots, Y_{n_2} \stackrel{\text{ind}}{\sim} \text{Normal}(\mu_2, \sigma_2^2)$ with sample variances S_1^2 and S_2^2, respectively. Then

$$\left(\frac{S_2^2}{S_1^2}F_{n_1-1,n_2-1,1-\alpha/2}, \frac{S_2^2}{S_1^2}F_{n_1-1,n_2-1,\alpha/2}\right)$$

  is a $(1-\alpha)100\%$ CI for σ_2^2/σ_1^2.

9. Sample size calculations

- Let $\hat{\theta}$ be an estimator of a parameter θ and let $\hat{\theta} \pm \eta$ be a CI for θ. Then the quantity η is called the *margin of error* (ME) of the CI.

- Let $X_1, \ldots, X_n$ be a random sample from a population with mean μ and variance σ^2. The smallest sample size n such that the ME of the CI

$$\bar{X}_n \pm z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

  is at most M^* is

$$n = \left\lceil \left(\frac{z_{\alpha/2}}{M^*}\right)^2 \sigma^2 \right\rceil,$$

  where $\lceil x \rceil$ is the smallest integer greater than or equal to x. Plug in an estimate $\hat{\sigma}^2$ of σ^2 from a previous study to make the calculation.

- Let $X_1, \ldots, X_n \stackrel{\text{ind}}{\sim} \text{Bernoulli}(p)$ distribution. The smallest sample size n such that the ME of the CI

$$\hat{p} \pm z_{\alpha/2}\sqrt{\frac{p(1-p)}{n}}$$

  is at most M^* is

$$n = \left\lceil \left(\frac{z_{\alpha/2}}{M^*}\right)^2 p(1-p) \right\rceil.$$

  Plug in an estimate $\hat{p}$ of p from a previous study to make the calculation or use $p = 1/2$ to err on the side of larger n.

- Let $X_1, \ldots, X_{n_1}$ be iid rvs with mean μ_1 and variance σ_1^2 and let $Y_1, \ldots, Y_{n_2}$ be iid rvs with mean μ_2 and variance σ_2^2. To find the smallest $n = n_1 + n_2$ such that the ME of the CI

$$\bar{X} - \bar{Y} \pm z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

is at most M^*, compute

$$n^* = \left(\frac{z_{\alpha/2}}{M^*}\right)^2 (\sigma_1 + \sigma_2)^2$$

and set

$$n_1 = \left\lceil \left(\frac{\sigma_1}{\sigma_1 + \sigma_2}\right) n^* \right\rceil \quad \text{and} \quad n_2 = \left\lceil \left(\frac{\sigma_2}{\sigma_1 + \sigma_2}\right) n^* \right\rceil.$$

Plug in estimates $\hat{\sigma}_1$ and $\hat{\sigma}_2$ for σ_1 and σ_2 from a previous study to make the calculation.

- Let $X_1, \ldots, X_{n_1} \overset{\text{ind}}{\sim} \text{Bernoulli}(p_1)$ and $Y_1, \ldots, Y_{n_2} \overset{\text{ind}}{\sim} \text{Bernoulli}(p_2)$ be independent random samples. To find the smallest $n = n_1 + n_2$ such that the ME of the CI

$$\hat{p}_1 - \hat{p}_2 \pm z_{\alpha/2}\sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}$$

is at most M^*, compute

$$n^* = \left(\frac{z_{\alpha/2}}{M^*}\right)^2 \left(\sqrt{p_1(1-p_1)} + \sqrt{p_2(1-p_2)}\right)^2.$$

Then set

$$n_1 = \left\lceil \left(\frac{\sqrt{p_1(1-p_1)}}{\sqrt{p_1(1-p_1)} + \sqrt{p_2(1-p_2)}}\right) n^* \right\rceil \quad \text{and} \quad n_2 = \left\lceil \left(\frac{\sqrt{p_2(1-p_2)}}{\sqrt{p_1(1-p_1)} + \sqrt{p_2(1-p_2)}}\right) n^* \right\rceil.$$

Plug in estimates $\hat{p}_1$ and $\hat{p}_2$ for p_1 and p_2 from a previous study to make the calculation.

10. First principles of estimation, sufficiency and Rao–Blackwell theorem

- Let $X_1, \ldots, X_n$ have a joint distribution depending on the parameter $\theta \in \Theta \subset \mathbb{R}^d$. A statistic $T(X_1, \ldots, X_n)$ is *a sufficient statistic for* θ if it carries all the information about θ contained in $X_1, \ldots, X_n$.
- Precisely, $T(X_1, \ldots, X_n)$ is *a sufficient statistic for* θ if the joint distribution of $X_1, \ldots, X_n$ conditional on the value of $T(X_1, \ldots, X_n)$ does not depend on θ.
  We can establish sufficiency using the following results:
  - For discrete rvs $(X_1, \ldots, X_n) \sim p_{X_1,\ldots,X_n}(x_1, \ldots, x_n; \theta)$, $T(X_1, \ldots, X_n)$ is a sufficient statistic for θ if for all $(x_1, \ldots, x_n)$ in the support of $(X_1, \ldots, X_n)$, the ratio

$$\frac{p_{(X_1,\ldots,X_n)}(x_1, \ldots, x_n; \theta)}{p_T(T(x_1, \ldots, x_n); \theta)}$$

    is free of θ, where $p_T(t; \theta)$ is the pmf of $T(X_1, \ldots, X_n)$.
  - For continuous rvs $(X_1, \ldots, X_n) \sim f_{X_1,\ldots,X_n}(x_1, \ldots, x_n; \theta)$, $T(X_1, \ldots, X_n)$ is a sufficient statistic for θ if for all $(x_1, \ldots, x_n)$ in the support of $(X_1, \ldots, X_n)$, the ratio

$$\frac{f_{X_1,\ldots,X_n}(x_1, \ldots, x_n; \theta)}{f_T(T(x_1, \ldots, x_n); \theta)}$$

    is free of θ, where $f_T(t; \theta)$ is the pdf of $T(X_1, \ldots, X_n)$.

- The *factorization theorem* gives another way to see whether a statistic is sufficient for θ:
  - For discrete rvs $(X_1, \dots, X_n) \sim p_{X_1,\dots,X_n}(x_1, \dots, x_n; \theta)$, $T(X_1, \dots, X_n)$ is a sufficient statistic for θ if and only if there exist functions $g(T; \theta)$ and $h(x_1, \dots, x_n)$ such that
    $$p_{X_1,\dots,X_n}(x_1, \dots, x_n; \theta) = g(T(x_1, \dots, x_n); \theta)h(x_1, \dots, x_n)$$
    for all $(x_1, \dots, x_n)$ in the support of $(X_1, \dots, X_n)$ and all $\theta \in \Theta$.
  - For continuous rvs $(X_1, \dots, X_n) \sim f_{X_1,\dots,X_n}(x_1, \dots, x_n; \theta)$, $T(X_1, \dots, X_n)$ is a sufficient statistic for θ if and only if there exist functions $g(T; \theta)$ and $h(x_1, \dots, x_n)$ such that
    $$f_{X_1,\dots,X_n}(x_1, \dots, x_n; \theta) = g(T(x_1, \dots, x_n); \theta)h(x_1, \dots, x_n)$$
    for all $(x_1, \dots, x_n)$ in the support of $(X_1, \dots, X_n)$ and all $\theta \in \Theta$.
- We also consider statistics with multiple components which take the form
  $$T(X_1, \dots, X_n) = (T_1(X_1, \dots, X_n), \dots, T_K(X_1, \dots, X_n))$$
  for some $K \geq 1$. For example, $T(X_1, \dots, X_n) = (X_{(1)}, \dots, X_{(n)})$ is the set of order statistics (which is always sufficient); here $K = n$.
- *Minimum-variance unbiased estimator (MVUE)*: Let $\hat{\theta}$ be an unbiased estimator of $\theta \in \Theta \subset \mathbb{R}$. If $\operatorname{Var} \hat{\theta} \leq \operatorname{Var} \tilde{\theta}$ for every unbiased estimator $\tilde{\theta}$ of θ, then $\hat{\theta}$ is a MVUE.
- *Rao–Blackwell Theorem:* Let $\tilde{\theta}$ be an estimator of $\theta \in \Theta \subset \mathbb{R}$ such that $\mathbb{E}\tilde{\theta} = \theta$ and let T be a sufficient statistic for θ. Define $\hat{\theta} = \mathbb{E}[\tilde{\theta}|T]$. Then $\hat{\theta}$ is an estimator of θ such that
  $$\mathbb{E}\hat{\theta} = \theta \quad \text{and} \quad \operatorname{Var} \hat{\theta} \leq \operatorname{Var} \tilde{\theta}.$$
  Interpretation: An unbiased estimator of θ can always be improved by taking into account the value of a sufficient statistic for θ. We conclude that a MVUE must be a function of a sufficient statistic.
- The MVUE for a parameter θ is (essentially) unique (for all situations considered in this course); to find it, do the following:
  (a) Find a sufficient statistic for θ.
  (b) Find a function of the sufficient statistic which is unbiased for θ. This is the MVUE.

  This also works for finding the MVUE of a function $\tau(\theta)$ of θ. In step (b), just find a function of the sufficient statistic which is unbiased for $\tau(\theta)$.

11. MoMs and MLEs

- Let $X_1, \dots, X_n$ be independent rvs with the same distribution as the rv X, where X has a distribution depending on the parameters $\theta_1, \dots, \theta_d$. If the first d moments $\mathbb{E}X, \dots, \mathbb{E}X^d$ of X are finite, then the *method of moments (MoMs) estimators* $\hat{\theta}_1, \dots, \hat{\theta}_d$ are the values of $\theta_1, \dots, \theta_d$ which solve the following system of equations:
  $$m'_1 := \frac{1}{n}\sum_{i=1}^{n} X_i = \mathbb{E}X =: \mu'_1(\theta_1, \dots, \theta_d)$$
  $$\vdots$$
  $$m'_d := \frac{1}{n}\sum_{i=1}^{n} X_i^d = \mathbb{E}X^d =: \mu'_d(\theta_1, \dots, \theta_d)$$

- Let $X_1, \ldots, X_n$ be a random sample from a distribution with pdf $f_X(x; \theta_1, \ldots, \theta_d)$ or pmf $p_X(x; \theta_1, \ldots, \theta_d)$ which depends on some parameters $(\theta_1, \ldots, \theta_d) \in \Theta \subset \mathbb{R}^d$. Then the *likelihood function* is defined as

$$\mathcal{L}(\theta_1, \ldots, \theta_d; X_1, \ldots, X_n) = \begin{cases} \prod_{i=1}^n f_X(X_i; \theta_1, \ldots, \theta_d), & \text{if } X_1, \ldots, X_n \overset{\text{ind}}{\sim} f_X(x; \theta_1, \ldots, \theta_d) \\ \prod_{i=1}^n p_X(X_i; \theta_1, \ldots, \theta_d), & \text{if } X_1, \ldots, X_n \overset{\text{ind}}{\sim} p_X(x; \theta_1, \ldots, \theta_d). \end{cases}$$

Moreover, the *log-likelihood function* is defined as

$$\ell(\theta_1, \ldots, \theta_d; X_1, \ldots, X_n) = \log \mathcal{L}(\theta_1, \ldots, \theta_d; X_1, \ldots, X_n).$$

- Let $X_1, \ldots, X_n$ be rvs with likelihood function $\mathcal{L}(\theta_1, \ldots, \theta_d; X_1, \ldots, X_n)$ for some parameters $(\theta_1, \ldots, \theta_d) \in \Theta \subset \mathbb{R}^d$. Then the *maximum likelihood estimators (MLEs)* of $\theta_1, \ldots, \theta_d$ are the values $\hat{\theta}_1, \ldots, \theta_d$ which maximize the likelihood function over all $(\theta_1, \ldots, \theta_d) \in \Theta$. That is

$$\begin{aligned} (\hat{\theta}_1, \ldots, \hat{\theta}_d) &= \underset{(\theta_1, \ldots, \theta_d) \in \Theta}{\operatorname{argmax}} \mathcal{L}(\theta_1, \ldots, \theta_d; X_1, \ldots, X_n) \\ &= \underset{(\theta_1, \ldots, \theta_d) \in \Theta}{\operatorname{argmax}} \ell(\theta_1, \ldots, \theta_d; X_1, \ldots, X_n) \end{aligned}$$

- If $\ell(\theta_1, \ldots, \theta_d; X_1, \ldots, X_n)$ is differentiable and has a single maximum in the interior of Θ, then we can find the MLEs $\hat{\theta}_1, \ldots, \theta_d$ by solving the following system of equations:

$$\begin{aligned} \frac{\partial}{\partial \theta_1} \ell(\theta_1, \ldots, \theta_d; X_1, \ldots, X_n) &= 0 \\ &\vdots \\ \frac{\partial}{\partial \theta_d} \ell(\theta_1, \ldots, \theta_d; X_1, \ldots, X_n) &= 0. \end{aligned}$$

- The MLE is always a function of a sufficient statistic.
- If $\hat{\theta}$ is the MLE for θ, then $\tau(\hat{\theta})$ is the MLE for $\tau(\theta)$, for any function τ.

pmf/pdf	$\mathcal{X}$	$M_X(t)$	$\mathbb{E}X$	$\operatorname{Var} X$
$p_X(x;p) = p^x(1-p)^{1-x}$,	$x = 0,1$	$pe^t + (1-p)$	p	$p(1-p)$
$p_X(x;n,p) = \binom{n}{x}p^x(1-p)^{n-x}$,	$x = 0,1,\ldots,n$	$[pe^t + (1-p)]^n$	np	$np(1-p)$
$p_X(x;p) = (1-p)^{x-1}p$,	$x = 1,2,\ldots$	$\frac{pe^t}{1-(1-p)e^t}$	p^{-1}	$(1-p)p^{-2}$
$p_X(x;p,r) = \binom{x-1}{r-1}(1-p)^{x-r}p^r$,	$x = r, r+1, \ldots$	$\left[\frac{pe^t}{1-(1-p)e^t}\right]^r$	rp^{-1}	$r(1-p)p^{-2}$
$p_X(x;\lambda) = e^{-\lambda}\lambda^x/x!$	$x = 0,1,\ldots$	$e^{\lambda(e^t-1)}$	λ	λ
$p_X(x;N,M,K) = \binom{M}{x}\binom{N-M}{K-x}/\binom{N}{K}$	$x = 0,1,\ldots,K$	¡complicadísimo!	$\frac{KM}{N}$	$\frac{KM}{N}\frac{(N-K)(N-M)}{N(N-1)}$
$p_X(x;K) = \frac{1}{K}$	$x = 1,\ldots,K$	$\frac{1}{K}\sum_{x=1}^K e^{tx}$	$\frac{K+1}{2}$	$\frac{(K+1)(K-1)}{12}$
$p_X(x;x_1,\ldots,x_n) = \frac{1}{n}$	$x = x_1,\ldots,x_n$	$\frac{1}{n}\sum_{i=1}^n e^{tx_i}$	$\bar{x} = \frac{1}{n}\sum_{i=1}^n x_i$	$\frac{1}{n}\sum_{i=1}^n (x_i - \bar{x})^2$
$f_X(x;\mu,\sigma^2) = \frac{1}{\sqrt{2\pi}}\frac{1}{\sigma}\exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$	$-\infty < x < \infty$	$e^{\mu t + \sigma^2 t^2/2}$	μ	σ^2
$f_X(x;\alpha,\beta) = \frac{1}{\Gamma(\alpha)\beta^\alpha}x^{\alpha-1}\exp\left(-\frac{x}{\beta}\right)$	$0 < x < \infty$	$(1-\beta t)^{-\alpha}$	$\alpha\beta$	$\alpha\beta^2$
$f_X(x;\lambda) = \frac{1}{\lambda}\exp\left(-\frac{x}{\lambda}\right)$	$0 < x < \infty$	$(1-\lambda t)^{-1}$	λ	λ^2
$f_X(x;\nu) = \frac{1}{\Gamma(\nu/2)2^{\nu/2}}x^{\nu/2-1}\exp\left(-\frac{x}{2}\right)$	$0 < x < \infty$	$(1-2t)^{-\nu/2}$	ν	2ν
$f_X(x;\alpha,\beta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}x^{\alpha-1}(1-x)^{\beta-1}$	$0 < x < 1$	$1 + \sum_{k=1}^\infty \frac{t^k}{k!}\left(\prod_{r=0}^k \frac{\alpha+r}{\alpha+\beta+r}\right)$	$\frac{\alpha}{\alpha+\beta}$	$\frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$

Table 1: Commonly encountered pmfs and pdfs along with their mgfs, expected values, and variances.

$$p_{(X_1,\ldots,X_K)}(x_1,\ldots,x_K;p_1,\ldots,p_K) = p_1^{x_1}\cdots p_K^{x_K}\cdot\mathbb{1}\left\{(x_1,\ldots,x_K)\in\{0,1\}^K : \textstyle\sum_{k=1}^K x_k = 1\right\}$$

$$p_{(Y_1,\ldots,Y_K)}(y_1,\ldots,y_K;n,p_1,\ldots,p_K) = \left(\frac{n!}{y_1!\cdots y_K!}\right)p_1^{y_1}\cdots p_K^{y_K}\cdot\mathbb{1}\left\{(y_1,\ldots,y_K)\in\{0,1,\ldots,n\}^K : \textstyle\sum_{k=1}^K y_k = n\right\}$$

$$f_{(X,Y)}(x,y;\mu_X,\mu_Y,\sigma_X^2,\sigma_Y^2,\rho) = \frac{1}{2\pi}\frac{1}{\sigma_X\sigma_Y\sqrt{1-\rho^2}}\exp\left(-\frac{1}{2}\frac{1}{1-\rho^2}\left[\left(\frac{x-\mu_X}{\sigma_X}\right)^2 - 2\rho\left(\frac{x-\mu_X}{\sigma_X}\right)\left(\frac{y-\mu_Y}{\sigma_Y}\right) + \left(\frac{y-\mu_Y}{\sigma_Y}\right)^2\right]\right)$$

Table 2: The "multinoulli" and multinomial pmfs and the bivariate Normal pdf.