Statistics 630 – Final Exam
Wednesday, 2 December 2020

Printed Name: ________________________ **Email:** ________________________

INSTRUCTIONS FOR THE STUDENT:

1. You have 2 hours to complete the exam (after taking a moment to read these instructions). Please indicate your start time: ______________________ ,

   and your end time: ______________________
2. There are 11 pages including this cover sheet and the formula sheets.
3. Questions 1–7 are multiple choice and worth 5 points each.
4. Questions 8–10 are multiple selection and worth 5 points each.
5. Questions 11–15 require solutions to be worked out and are 10 points each. Please write out your answers in *the spaces provided*, explaining your steps. You may refer to theorems by name/description rather than by its number in the book.
6. If you *cannot* print out the exam, please write your answers on blank sheet of paper – in order.
7. You may use the *attached formula sheets*. No other resources are allowed. Do not use the textbook, the class notes, homework or formula sheets that were posted online.
8. *Do not use a calculator.* You may leave answers in forms that can easily be put into a calculator such as $\frac{12}{19}$, $\binom{40}{5}$, e^{-3}, $\Phi(1.5)$, etc.
9. Do not discuss or provide any information to anyone concerning any of the questions on this exam until your solutions are returned or I post my solutions.

I attest that I spent no more than 2 hours to complete the exam. I used only the materials allowed above. I did not receive assistance from or provide assistance to anyone either before or while taking this exam.

Student's Signature ____________________________________

Questions 1–7 are multiple choice: circle the single correct answer. No partial credit!

1. (5 points) $\mathsf{Var}(X) = 4$, $\mathsf{Var}(Y) = 1$ and $\mathsf{Cov}(X, Y) = -2$. What is $\mathsf{Var}(X - Y)$?

   (a) 1.
   (b) 9.
   (c) 3.
   (d) 0.
   (e) 5.

2. (5 points) Sally Scientist obtained a 98% confidence interval for μ from a random sample of normal$(\mu, 37.0)$ data (σ^2 known). The sample size was $n = 25$ and the interval she computed was $9.54 \pm 2.83 = (6.71, 12.37)$. The confidence level 98% is the probability, before sampling,

   (a) that $\bar{X}$ will be between 6.71 and 12.37.
   (b) that μ is between 6.71 and 12.37.
   (c) that $\bar{X}$ will be within 2.83 of 9.54.
   (d) that $\bar{X}$ will be within 2.83 of μ.
   (e) that a Type I error is not made.

3. (5 points) T is a statistic with cdf $F(t) = \theta t + (1 - \theta)t^2$, $0 \leq t \leq 1$, $0 \leq \theta \leq 1$. Suppose $H_0 : \theta = 0$ is rejected when $T \leq \sqrt{\alpha}$, which is a size α test. Then the power function for the test is

   (a) $\theta T + (1 - \theta)T^2$.
   (b) $\theta\sqrt{\alpha} + (1 - \theta)\alpha$.
   (c) $\Phi\left(\frac{\sqrt{\alpha} - .5}{\sqrt{1/12}}\right)$.
   (d) $\theta\alpha + (1 - \theta)\alpha^2$.
   (e) $\Phi\left(\frac{\sqrt{\alpha} - \theta}{\sqrt{1/12}}\right)$.

4. (5 points) An estimator $\widehat{\theta}$ is asymptotically normal$(\theta, \frac{1}{n\theta})$. Consequently, a level γ Wald confidence interval for θ is

   (a) $\widehat{\theta} \pm z_{(1+\gamma)/2} \dfrac{1}{\sqrt{n\widehat{\theta}}}$.
   (b) $\widehat{\theta} \pm \sqrt{\dfrac{\chi^2_\gamma}{n\theta}}$.
   (c) $\widehat{\theta} \pm z_{(1+\gamma)/2} \dfrac{\widehat{\theta}}{\sqrt{n}}$.
   (d) $\widehat{\theta} \pm z_{(1+\gamma)/2} \dfrac{1}{\sqrt{n\theta}}$.
   (e) $\widehat{\theta} \pm \sqrt{\dfrac{z_{(1+\gamma)/2}}{n\widehat{\theta}}}$.

5. (5 points) A and B are <u>independent</u> events with probabilities α and β, respectively. The conditional probability that both occur, given at least one occurs, (that is, $\mathsf{P}(A \cap B \mid A \cup B)$) is

   (a) $\frac{\alpha\beta}{\alpha+\beta}$.

   (b) $\frac{\alpha+\beta}{\alpha\beta}$.

   (c) $\frac{\alpha\beta}{(1-\alpha)(1-\beta)}$.

   (d) $\frac{\alpha+\beta}{2-\alpha-\beta}$.

   (e) $\frac{\alpha\beta}{\alpha+\beta-\alpha\beta}$.

6. (5 points) In the spirit of diversity, independent random variables $X \sim \text{normal}(1, 2)$, $Y \sim \text{Poisson}(3)$ and $Z \sim \text{gamma}(4, 5)$ get together for a (socially distanced) party. Let $T = X + Y + Z$ be their sum. The variance of T is

   (a) $2 \times 3 \times 5$.

   (b) $2 + 3 + 5$.

   (c) $2 + 3 + 20$.

   (d) $2 + 3 + \frac{4}{5}$.

   (e) $2 + 3 + \frac{4}{25}$.

7. (5 points) Consider the basic two-sided test for a mean (with variance known) having rejection criterion $\sqrt{n}\left|\frac{\bar{X}-\mu_0}{\sigma}\right| > z_{1-\alpha/2}$. The power of the test increases

   (a) when α decreases or n increases.

   (b) only when both α decreases and n increases.

   (c) when α increases or n increases.

   (d) only when n increases.

   (e) only when α decreases.

Questions 8–10 can have 0, 1 or more than 1 answer: circle <u>all</u> that are correct.

8. (5 points) $X_1, \ldots, X_n$ are iid normal(2,1) random variables. Which of the following hold?

   (a) $(X_1 - 2)^2 \sim \text{chi-square}(1)$.

   (b) $\frac{(X_1-2)^2}{((X_1-2)^2+\cdots+(X_n-2)^2)/n} \sim \mathrm{F}(1, n)$.

   (c) $\frac{(X_1-2)^2}{(X_2-2)^2+\cdots+(X_n-2)^2} \sim \mathrm{F}(1, n-1)$.

   (d) $\sum_{i=1}^{n}(X_i - \bar{X})^2 \sim \text{chi-square}(n-1)$.

   (e) $\sum_{i=1}^{n} X_i^2 \sim \text{chi-square}(n)$.

9. (5 points) $Q_1, \ldots, Q_n$ is a random sample from a distribution with mean $\sqrt{\beta}$ and variance 2β. Which of the following are correct?

   (a) $\bar{Q}$ is unbiased for $\sqrt{\beta}$.

   (b) $\frac{1}{n}\sum_{i=1}^{n} Q_i^2$ is unbiased for 2β.

   (c) $\bar{Q}^2$ is unbiased for β.

   (d) $\frac{\bar{Q}^2}{1+2/n}$ is unbiased for β.

   (e) $\frac{1}{2(n-1)}\sum_{i=1}^{n}(Q_i - \bar{Q})^2$ is unbiased for β.

10. (5 points) A Poisson(λ) random sample $N_1, \ldots, N_n$ has MLE $\widehat{\lambda} = \bar{N}$, score $S(\lambda) = n(\frac{\bar{N}}{\lambda} - 1)$ and Fisher information $I(\lambda) = \frac{n}{\lambda}$. Which of the following hold, as $n \to \infty$?

   (a) $\frac{\bar{N}-\lambda}{\sqrt{n/\lambda}} \xrightarrow{\mathsf{D}}$ normal$(0, 1)$.

   (b) $\frac{(\bar{N}-\lambda)^2}{\lambda/n} \xrightarrow{\mathsf{D}}$ chi-square(1).

   (c) $\frac{S^2(\lambda)}{I(\lambda)} \xrightarrow{\mathsf{D}}$ chi-square(n).

   (d) $\bar{N} \xrightarrow{\mathsf{D}}$ normal$(\lambda, \lambda/n)$.

   (e) $\frac{S(\lambda)}{\sqrt{I(\lambda)}} \xrightarrow{\mathsf{D}}$ normal$(0, 1)$.

Provide solutions to Questions 11–15, to the point of a calculable expression.

11. X has pdf $f_X(x) = \dfrac{e^{-x}}{c(1+x)}$, $x > 0$ (where $c = \int_0^\infty \frac{e^{-x}}{1+x}\,dx$), and the conditional pdf for Y, given $X = x$, is $f_{Y|X}(y|x) = (1+x)e^{-y-xy}$, $y > 0$. Find the conditional pdf for X, given $Y = y$.

12. (10 points) $R_1, R_2, \ldots$ are iid random variables from the distribution with pdf $f(r) = 2\theta r \mathrm{e}^{-\theta r^2}$, $r > 0$, $\theta > 0$. Determine the MLE for θ and its asymptotic variance.

13. (10 points) Assume as in the previous problem but suppose θ has prior distribution gamma(2,1) with pdf $\theta \mathrm{e}^{-\theta}$. Identify the posterior distribution for θ (by name and parameter values).

14. (10 points) Consider an exponential(λ) random sample $X_1, \ldots, X_n$. Determine a moment estimator for $\mathrm{e}^{-\lambda} = \mathrm{e}^{-1/\mu}$, based on $\bar{X}$, and use the delta method to derive an asymptotic variance for it.

15. (10 points) Let $Y_1, Y_2, \ldots, Y_n$ be a random sample from the geometric(θ) distribution with pmf $\theta(1-\theta)^y$, $y = 0, 1, 2 \ldots$, $0 < \theta < 1$. Determine the size α (asymptotic) generalized likelihood ratio test for $H_0 : \theta = \frac{1}{2}$ versus $H_a : \theta \neq \frac{1}{2}$.

Formulas for Final Exam

Bayes' rule $\mathsf{P}(B_j \mid A) = \frac{\mathsf{P}(A|B_j)\mathsf{P}(B_j)}{\sum_{k=1}^n \mathsf{P}(A|B_k)\mathsf{P}(B_k)}$ if $B_1, \ldots, B_n$ are disjoint and $\bigcup_{k=1}^n B_k = S$.

quantile function $Q_X(p)$ satisfies $F_X(x) \le p \le F(Q_X(p))$ for all $x < Q_X(p)$. $F(Q_X(p)) = p$ if X is a continuous rv.

distribution of a function of X $F_Y(y) = \mathsf{P}(h(X) \le y)$ for $Y = h(X)$.

If X is a discrete rv or $h(x)$ takes only countably many values then Y has pmf $p_Y(y) = \mathsf{P}(h(X) = y)$.

If X is a continuous rv and $h(x)$ is a continuous function then Y has pdf $f_Y(y) = \frac{\mathrm{d}x}{\mathrm{d}y}\,\mathsf{P}(h(X) \le y)$.

binomial theorem $\sum_{k=0}^n \binom{n}{k} a^k b^{n-k} = (a+b)^n$.

geometric sum $\sum_{k=n}^\infty a^k = \frac{a^n}{1-a}$ if $-1 < a < 1$.

exponential expansion $\sum_{k=0}^\infty \frac{a^k}{k!} = \mathrm{e}^a$.

gamma integral $\int_0^\infty x^a \mathrm{e}^{-x}\,\mathrm{d}x = \Gamma(a+1) = a!$ for $a > -1$.

Bernoulli pmf $p(x) = (1-\theta)^{1-x}\theta^x I_{\{0,1\}}(x)$ for $0 < \theta < 1$, same as binomial$(1, \theta)$.

beta(a, b) pdf $f(x) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\, x^{a-1}(1-x)^{b-1} I_{(0,1)}(x)$ for $a > 0$, $b > 0$; $\mathsf{E}(X) = \frac{a}{a+b}$ $\mathsf{Var}(X) = \frac{ab}{(a+b)^2(a+b+1)}$.

binomial(n, θ) pmf $p(x) = \binom{n}{x}\theta^x(1-\theta)^{n-x} I_{\{0,1,\ldots,n\}}(x)$ for $0 < \theta < 1$. $\mathsf{E}(X) = n\theta$, $\mathsf{Var}(X) = n\theta(1-\theta)$, $m(s) = (1 - \theta + \theta\mathrm{e}^s)^n$.

chi-square(n) same as gamma$(\frac{n}{2}, \frac{1}{2})$, the distribution of $X = Z_1^2 + \cdots + Z_n^2$ for iid standard normal $Z_1, \ldots, Z_n$. $\mathsf{E}(X) = n$, $\mathsf{Var}(X) = 2n$.

In particular, if $X_1, \ldots, X_n \overset{\text{iid}}{\sim}$ normal(μ, σ^2) then $\frac{(n-1)S^2}{\sigma^2} \sim$ chi-square$(n-1)$.

discrete uniform(N) pmf $p(x) = \frac{1}{N}\, I_{\{1,2,\ldots,N\}}(x)$. $\mathsf{E}(X) = \frac{N+1}{2}$, $\mathsf{Var}(X) = \frac{N^2-1}{12}$.

exponential(λ) pdf $f(x) = \lambda \mathrm{e}^{-\lambda x} I_{(0,\infty)}(x)$ for $\lambda > 0$, same as gamma$(1, \lambda)$. $\mathsf{E}(X) = \frac{1}{\lambda}$, $\mathsf{Var}(X) = \frac{1}{\lambda^2}$.

F(m, n) the distribution of $W = \frac{X/m}{Y/n}$ where $X \sim$ chi-square(m), $Y \sim$ chi-square(n), independent. $\mathsf{E}(W) = \frac{n}{n-2}$ if $n > 2$.

gamma(α, λ) pdf $f(x) = \frac{\lambda^\alpha}{\Gamma(\alpha)}\, x^{\alpha-1}\mathrm{e}^{-\lambda x} I_{(0,\infty)}(x)$ for $\lambda > 0$, $\alpha > 0$; $\mathsf{E}(X) = \frac{\alpha}{\lambda}$, $\mathsf{Var}(X) = \frac{\alpha}{\lambda^2}$, $m(s) = \left(\frac{\lambda}{\lambda - s}\right)^\alpha$ if $s < \lambda$.

geometric(θ) pmf $p(x) = \theta(1-\theta)^x I_{\{0,1,2,\ldots\}}(x)$ for $0 < \theta < 1$, same as negative binomial$(1, \theta)$. $\mathsf{E}(X) = \frac{1-\theta}{\theta}$, $\mathsf{Var}(X) = \frac{1-\theta}{\theta^2}$.

hypergeometric(N, M, n) pmf $p(x) = \frac{\binom{M}{x}\binom{N-M}{n-x}}{\binom{N}{n}} I_{\{0,1,\ldots,n\}}(x)$ for $M < N$. $\mathsf{E}(X) = np$ where $p = \frac{M}{N}$, $\mathsf{Var}(X) = \frac{N-n}{N-1}\, np(1-p)$.

negative binomial(r, θ) pmf $p(x) = \binom{r+x-1}{r-1}\theta^r(1-\theta)^x I_{\{0,1,2,\ldots\}}(x)$ for $0 < \theta < 1$. $\mathsf{E}(X) = \frac{r(1-\theta)}{\theta}$, $\mathsf{Var}(X) = \frac{r(1-\theta)}{\theta^2}$, $m(s) = \left(\frac{\theta}{1-(1-\theta)\mathrm{e}^s}\right)^r$ if $s < -\log(1-\theta)$.

normal(μ, σ^2) pdf $f(x) = \frac{1}{\sqrt{2\pi\sigma^2}}\, \mathrm{e}^{-(x-\mu)^2/(2\sigma^2)} I_{(-\infty,\infty)}(x)$ for $\sigma^2 > 0$; $\mathsf{E}(X) = \mu$, $\mathsf{Var}(X) = \sigma^2$, $m(s) = \mathrm{e}^{\mu s + \sigma^2 s^2/2}$.

Poisson(λ) pmf $p(x) = \frac{\lambda^x}{x!}\, \mathrm{e}^{-\lambda} I_{\{0,1,2,\ldots\}}(x)$ for $\lambda > 0$. $\mathsf{E}(X) = \lambda$, $\mathsf{Var}(X) = \lambda$, $m(s) = \mathrm{e}^{\lambda(\mathrm{e}^s - 1)}$.

t(n) the distribution of $T = \frac{Z}{\sqrt{Y/n}}$ where $Z \sim \text{normal}(0,1)$, $Y \sim \text{chi-square}(n)$, independent. $\mathsf{E}(T) = 0$, $\mathsf{Var}(T) = \frac{n}{n-2}$ if $n > 2$. In particular, if $X_1, \ldots, X_n \overset{\text{iid}}{\sim} \text{normal}(\mu, \sigma^2)$ then $\frac{\bar{X}-\mu}{S/\sqrt{n}} \sim \text{t}(n-1)$.

uniform(a, b) pdf $f(x) = \frac{1}{b-a}\, I_{(a,b)}(x)$ for $a < b$. $\mathsf{E}(X) = \frac{a+b}{2}$, $\mathsf{Var}(X) = \frac{(b-a)^2}{12}$.

Weibull(α, β) pdf $f(x) = \frac{\alpha}{\beta}\, (x/\beta)^{\alpha-1} \mathrm{e}^{-(x/\beta)^\alpha} I_{(0,\infty)}(x)$ for $\alpha > 0$, $\beta > 0$. $\mathsf{E}(X^k) = \beta^k\, \Gamma(1 + \frac{k}{\alpha})$.

marginal pmf/pdf $p_X(x) = \sum_y p(x,y)$; $f_X(x) = \int_{-\infty}^{\infty} f(x,y)\, dy$.

conditional pmf/pdf $p_{X|Y}(x|y) = \frac{p(x,y)}{p_Y(y)}$; $f_{X|Y}(x|y) = \frac{f(x,y)}{f_Y(y)}$.

independent random variables $p(x,y) = p_X(x)p_Y(y)$ if (X,Y) is discrete; $f(x,y) = f_X(x)f_Y(y)$ if (X,Y) is continuous.

discrete convolution $p_{X+Y}(z) = \sum_x p_X(x)p_Y(z-x)$ for independent X, Y.

continuous convolution $f_{X+Y}(z) = \int_{-\infty}^{\infty} f_X(x)f_Y(z-x)\, dx$ for independent X, Y.

covariance and correlation $\mathsf{Cov}(X,Y) = \mathsf{E}((X-\mu_X)(Y-\mu_Y)) = \mathsf{E}(XY) - \mu_X\mu_Y$; $\mathsf{Corr}(X,Y) = \frac{\mathsf{Cov}(X,Y)}{\sigma_X\sigma_Y}$. For independent X and Y, $\mathsf{Cov}(X,Y) = \mathsf{Corr}(X,Y) = 0$.

expectation of a sum $\mathsf{E}(a_1X_1 + \cdots + a_nX_n) = a_1\mathsf{E}(X_1) + \cdots + a_n\mathsf{E}(X_n)$.

expectation of a product If $X_1, \ldots, X_n$ are independent, $\mathsf{E}\left(\prod_{i=1}^n h_i(X_i)\right) = \prod_{i=1}^n \mathsf{E}(h_i(X_i))$.

variance of a sum $\mathsf{Var}(aX + bY) = a^2\,\mathsf{Var}(X) + 2ab\,\mathsf{Cov}(X,Y) + b^2\,\mathsf{Var}(Y)$.

variance of a sum of independent rvs $\mathsf{Var}(a_1X_1 + \cdots + a_nX_n) = a_1^2\,\mathsf{Var}(X_1) + \cdots + a_n^2\,\mathsf{Var}(X_n)$.

moments k-th moment is $\mu_k = \mathsf{E}(X^k)$, $k = 1, 2, \ldots$

moment generating function $m_X(s) = \mathsf{E}(\mathrm{e}^{sX})$; $\mathsf{E}(X^k) = \frac{\mathrm{d}x^k}{\mathrm{d}s^k}\, m_X(s)\Big|_{s=0}$.

mgf of a sum If X and Y are independent, $m_{aX+bY}(s) = \mathsf{E}(\mathrm{e}^{(aX+bY)s}) = m_X(as)m_Y(bs)$.

conditional expectation $\mathsf{E}(h(Y)|X=x) = \sum_y h(y)p_{Y|X}(y|x)$ or $\mathsf{E}(h(Y) \mid X = x) = \int_{-\infty}^{\infty} h(y)f_{Y|X}(y|x)\, dy$.

iterated expectation $\mathsf{E}(h(Y)) = \mathsf{E}(\mathsf{E}(h(Y) \mid X))$, $\mathsf{E}(g(X)h(Y)) = \mathsf{E}(g(X)\mathsf{E}(h(Y) \mid X))$.

conditional variance $\mathsf{Var}(Y \mid X) = \mathsf{E}(Y^2 \mid X) - (\mathsf{E}(Y \mid X))^2$.

variance partition formula $\mathsf{Var}(Y) = \mathsf{E}(\mathsf{Var}(Y \mid X)) + \mathsf{Var}(\mathsf{E}(Y \mid X))$.

Markov's inequality $\mathsf{P}(|X| \geq x) \leq \frac{\mathsf{E}(|X|)}{x}$ for $x > 0$.

Chebyshev's inequality $\mathsf{P}(|X - \mu_X| \geq x) \leq \frac{\mathsf{Var}(X)}{x^2}$ for $x > 0$.

sample mean, variance, k-th moment $\bar{X} = \frac{1}{n}\sum_{i=1}^n X_i$; $\widehat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^n (X_i - \bar{X})^2$; $m_k = \frac{1}{n}\sum_{i=1}^n X_i^k$.

unbiased sample variance $S^2 = \frac{n}{n-1}\widehat{\sigma}^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \bar{X})^2$.

law of large numbers For iid $X_1, X_2, \ldots$ with mean μ, $\bar{X}_n \to \mu$ as $n \to \infty$.

central limit theorem For iid $X_1, X_2, \ldots$ with mean μ and variance σ^2, $\mathsf{P}\Big(\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \leq z\Big) = \mathsf{P}\Big(\frac{X_1 + \cdots + X_n - n\mu}{\sqrt{n}\sigma} \leq z\Big) \to \Phi(z)$ (normal(0,1) cdf), as $n \to \infty$.

bias and standard error $\mathsf{Bias}(\widehat{\theta}) = \mathsf{E}(\widehat{\theta}) - \theta$; $\mathsf{SE}(\widehat{\theta}) = \sqrt{\mathsf{Var}(\widehat{\theta})}$.

mean squared error $\mathsf{MSE}(\widehat{\theta}) = \mathsf{E}((\widehat{\theta} - \theta)^2) = \mathsf{Var}(\widehat{\theta}) + (\mathsf{Bias}(\widehat{\theta}))^2$.

consistency $\widehat{\theta}$ is consistent if $\mathsf{MSE}(\widehat{\theta}) \to 0$ as $n \to \infty$.

method of moments for iid sample match the k-th population moment $\mathsf{E}(X^k)$ with the k-th sample moment m_k, and solve for the desired parameter estimates.

likelihood function $L(\theta|X_1, \ldots, X_n) = \prod_{i=1}^n f_\theta(X_i)$ for iid sample $\underline{X} = (X_1, \ldots, X_n)$.

maximum likelihood for iid sample maximize the likelihood function $L(\theta|X_1, \ldots, X_n) = \prod_{i=1}^n f_\theta(X_i)$ or the log-likelihood $\ell(\theta|X_1, \ldots, X_n) = \log L(\theta|X_1, \ldots, X_n) = \sum_{i=1}^n \log f_\theta(X_i)$.

If $\log L(\theta)$ is differentiable and concave at θ, the MLE is a solution to $S(\theta) = \frac{d}{d\theta}\log L(\theta) = 0$. (For a multidimensional parameter θ this is a system of equations.)

score function $S(\theta|X_1, \ldots, X_n) = \frac{\mathrm{d}}{\mathrm{d}\theta}\ell(\theta)$.

Fisher information $I_n(\theta) = \mathsf{Var}(\frac{\mathrm{d}}{\mathrm{d}\theta}\ell(\theta)) = -\mathsf{E}(\frac{\mathrm{d}^2}{\mathrm{d}\theta^2}\ell(\theta))$, if ℓ has two derivatives.

For an iid sample, $I_n(\theta) = nI_1(\theta)$ and $I_1(\theta) = \mathsf{Var}(\frac{\mathrm{d}}{\mathrm{d}\theta}\log f_\theta(X_1)) = -\mathsf{E}(\frac{\mathrm{d}^2}{\mathrm{d}\theta^2}\log f_\theta(X_1))$.

sufficient statistic $T = T(\underline{X})$ is sufficient if $L(\theta|\underline{X}) = h(\underline{X})g(T(\underline{X}), \theta)$ for some functions $h(\underline{x})$ and $g(t, \theta)$.

exponential family The pdf/pmf has the form $f_X(x|\theta) = d(\theta)h(x)\mathrm{e}^{c(\theta)t(x)}$ for all x, θ. In this case, with an iid random sample, $T(\underline{X}) = \sum_{i=1}^n t(X_i)$ is a sufficient statistic and $I_n(\theta) = n(c'(\theta))^2\,\mathsf{Var}(t(X_1))$.

asymptotics for MLE Assuming Fisher information exists and $\widehat{\theta}$ is the MLE, $\widehat{\theta} \to \theta$ in probability and $\sqrt{I_n(\theta)}(\widehat{\theta} - \theta) \to \text{normal}(0, 1)$ in distribution as $n \to \infty$.

asymptotic normality $\widehat{\theta}$ is asymptotic normal(θ, V_n) if $\frac{\widehat{\theta} - \theta}{\sqrt{V_n}} \to \text{normal}(0, 1)$ in distribution as $n \to \infty$. V_n may depend on θ or other parameters. If $\widehat{V}_n$ is an estimator such that $\widehat{V}_n/V_n \to 1$ then $\frac{\widehat{\theta} - \theta}{\sqrt{\widehat{V}_n}} \to \text{normal}(0, 1)$ in distribution.

delta method If $g(\theta)$ is continuously differentiable and estimator $\widehat{\theta}$ is asymptotic normal(θ, V_n), then $g(\widehat{\theta})$ is asymptotic normal$(g(\theta), (g'(\theta))^2 V_n)$.

level γ confidence interval $(L(\underline{\mathrm{X}}), U(\underline{\mathrm{X}}))$ such that $\mathsf{P}_\theta(L(\underline{\mathrm{X}}) \le \theta \le U(\underline{\mathrm{X}})) = \gamma$.

confidence interval from pivot If $h(\underline{\mathrm{X}}, \theta)$ has a distribution that does not depend on θ, a level γ confidence interval is defined by $\{\theta : h(\underline{\mathrm{X}}, \theta) \in A\}$ where $\mathsf{P}_\theta(h(\underline{\mathrm{X}}, \theta) \in A) = \gamma$.

Wald confidence interval If $\widehat{\theta}$ is asymptotic normal(θ, V_n) and $\widehat{V}_n$ is an estimator for V_n, an approximate level γ confidence interval for θ has endpoints $\widehat{\theta} \pm z_{(1+\gamma)/2}\sqrt{\widehat{V}_n}$, where $z_{(1+\gamma)/2}$ is the $(1+\gamma)/2$ quantile of the normal(0,1) distribution.

score confidence interval For MLE $\widehat{\theta}$, an approximate level γ confidence interval defined by $\{\theta : -z_{(1+\gamma)/2} \le (\frac{\mathrm{d}}{\mathrm{d}\theta}\,\ell(\theta))/\sqrt{I_n(\theta)} \le z_{(1+\gamma)/2}\}$, where $z_{(1+\gamma)/2}$ is the $(1+\gamma)/2$ quantile of the normal(0,1) distribution.

A related method is the interval given by $\{\theta : -z_{(1+\gamma)/2} \le \sqrt{I_n(\theta)}(\widehat{\theta} - \theta) \le z_{(1+\gamma)/2}\}$.

Type I and II errors, level and power A Type I error is rejecting H_0 when it is true. The level of a test is $\alpha = \max_{\theta \in H_0} \mathsf{P}_\theta(H_0 \text{ is rejected})$ computed with values of θ such that H_0 true.

A Type II error is not rejecting H_0 when H_a is true. The power of a test is $\beta = \beta(\theta) = \mathsf{P}_\theta(H_0 \text{ is rejected})$ computed with parameter value θ (satisfying H_a).

P-value The smallest level α for which H_0 will still be rejected – it is a statistic (function of the data).

Neyman-Pearson likelihood ratio test For simple hypotheses $H_0 : \theta = \theta_0$ vs. $H_a : \theta = \theta_1$, reject H_0 if $LR = \frac{L(\theta_1)}{L(\theta_0)} \ge c_\alpha$ where $\mathsf{P}(LR \ge c_\alpha) = \alpha$ when H_0 is true.
If, for each c, $LR \ge c \iff T \ge k$ (or $R \ge c \iff T \le k$) for some statistic T and some value k then it suffices to find k_α such that $\mathsf{P}(T \ge k_\alpha) = \alpha$ (resp., $\mathsf{P}(T \le k_\alpha) = \alpha$) when H_0 is true.

generalized likelihood ratio test For hypotheses H_0 and H_a about parameter θ and MLE $\widehat{\theta}$, reject H_0 if $LR = \frac{L(\widehat{\theta})}{\max_{\theta \in H_0} L(\theta)} \ge c_\alpha$ where $\max_{\theta \in H_0} \mathsf{P}(LR \ge c_\alpha) = \alpha$.

If $H_0 : \theta = \theta_0$ and $H_a : \theta \ne \theta_0$ then $LR = \frac{L(\widehat{\theta})}{L(\theta_0)}$.

uniformly most powerful test A test is UMP if it has maximum possible power for every parameter value θ that satisfies H_a.
In particular, if the test is the same as the Neyman-Pearson test for each θ satisfying H_a then it is UMP.

Wald test If $\widehat{\theta}$ is asymptotic normal(θ, V_n) and $\widehat{V}_n$ is an estimator for V_n, reject $H_0 : \theta = \theta_0$ when $\frac{|\widehat{\theta} - \theta_0|}{\sqrt{\widehat{V}_n}} \ge z_{1-\alpha/2}$, where $z_{1-\alpha/2}$ is the $(1-\alpha/2)$ quantile of the normal(0,1) dist. Equivalently, reject H_0 if $\frac{(\widehat{\theta} - \theta_0)^2}{\widehat{V}_n} \ge \chi^2_{1,1-\alpha}$.

Important case: $\widehat{\theta}$ is the MLE and $\widehat{V}_n = 1/I_n(\widehat{\theta})$.

(asymptotic) score test For MLE $\widehat{\theta}$, reject $H_0 : \theta = \theta_0$ when $\frac{|\frac{\mathrm{d}}{\mathrm{d}\theta}\,\ell(\theta_0)|}{\sqrt{I_n(\theta_0)}} \geq z_{1-\alpha/2}$, where $z_{1-\alpha/2}$ is the $(1-\alpha/2)$ quantile of the normal(0,1) dist. Equivalently, reject H_0 if $\frac{(\frac{\mathrm{d}}{\mathrm{d}\theta}\,\ell(\theta_0))^2}{I_n(\theta_0)} \geq \chi^2_{1,1-\alpha}$.

A related test is to reject $H_0 : \theta = \theta_0$ when $\sqrt{I_n(\theta_0)}\,|\widehat{\theta} - \theta_0| \geq z_{1-\alpha/2}$, and $\widehat{\theta}$ is the MLE.

asymptotic likelihood ratio test Using the generalized LR statistic, reject $H_0 : \theta = \theta_0$ when $2\log(LR) \geq z^2_{1-\alpha/2} = \chi^2_{1,1-\alpha}$.

test equivalent to interval Define a test from an interval (or an interval from a test) by: reject $H_0 : \theta = \theta_0$ at level $\alpha \iff \theta_0$ is not in the $1-\alpha$ confidence interval.

prior and posterior distributions If the prior density (or pmf) for θ is $f_\Theta(\theta)$ then the posterior density (or pmf) is $f_\Theta(\theta|\underline{\mathrm{X}}) = c(\underline{\mathrm{X}})f_{\underline{\mathrm{X}}}(\underline{\mathrm{X}}|\theta)f_\Theta(\theta)$, with $c(\underline{\mathrm{X}})$ chosen so that $f_\Theta(\theta|\underline{\mathrm{X}})$ is a proper pdf (pmf) in θ.

Bayes estimator Either the mean or the mode of the posterior distribution.

Bayes γ credible interval An interval $(L(\underline{\mathrm{X}}), U(\underline{\mathrm{X}}))$ such that, under the posterior distribution, $\mathsf{P}(L(\underline{\mathrm{X}}) \leq \theta \leq U(\underline{\mathrm{X}}) \mid \underline{\mathrm{X}}) = \gamma$.
The interval is HPD (highest posterior density) if it equals the set $\{\theta : f_\Theta(\theta|\underline{\mathrm{X}}) \geq c\}$ for some constant c.

Bayes Hypothesis test Choose H_1 if and only if $\dfrac{\mathsf{P}(H_1 \mid \underline{\mathrm{X}})}{\mathsf{P}(H_0 \mid \underline{\mathrm{X}})} > 1$.