Math 404 Exam 2 is Wed. March 21. **You are allowed 14 sheets of notes and a calculator.** The exam covers HW4-6, and Q4-6.

47) Often $L(\boldsymbol{\theta}) = L(\boldsymbol{\theta}|x_1, ..., x_n) = \prod_{i=1}^n L(\boldsymbol{\theta}|x_i)$. Note that $1 - F(w) = S(w)$.

a) For iid individual data, $L(\boldsymbol{\theta}|x_i) = f(x_i)$ if X has pdf $f(x)$.

b) For iid individual data, $L(\boldsymbol{\theta}|x_i) = p(x_i)$ if X has pmf $p(x)$.

c) If it is only known that x_i is in some interval $(c_{j-1}, c_j]$, then
$L(\boldsymbol{\theta}|x_i) = P(x_i \in (c_{j-1}, c_j]) = F(c_j) - F(c_{j-1})$.
The endpoints can be open or closed if X is from a continuous distribution.

d) If x_i is right censored at u_i, then the interval is $[u_i, \infty)$, and $L(\boldsymbol{\theta}|x_i) = 1 - F(u_i)$.

e) For grouped data from the table below, $L(\boldsymbol{\theta}) = \prod_{j=1}^{m} [F(c_j) - F(c_{j-1})]^{n_j}$.

interval	number
$(c_0, c_1]$	n_1
$(c_1, c_2]$	n_2
$(c_2, c_3]$	n_3
$\vdots$	$\vdots$
$(c_{m-2}, c_{m-1}]$	n_{m-1}
$(c_{m-1}, c_m]$	n_m

f) If x_i is left truncated at d_i, then $L(\boldsymbol{\theta}|x_i) = \dfrac{f(x_i)}{1 - F(d_i)}$.

g) If x_i is left truncated at d_i and right censored at u_i, then $L(\boldsymbol{\theta}|x_i) = \dfrac{1 - F(u_i)}{1 - F(d_i)}$.

h) If the data are left truncated at the deductible d with $n - k$ uncensored cases x_i and k cases right censored at u, then $L(\boldsymbol{\theta}) = \dfrac{[\prod_{i=1}^{n-k} f(x_i)][1 - F(u)]^k}{[1 - F(d)]^n}$.

i) (**Rare**, the interval is $(0, d]$.): If x_i is censored below at d, $L(\boldsymbol{\theta}|x_i) = F(d)$.

j) (**Rare**): If x_i is truncated above at u, $L(\boldsymbol{\theta}|x_i) = \dfrac{f(x_i)}{F(u)}$.

Note that left truncated = truncated below = truncated, and right censored = censored above = censored are often used.

48) Let $w_i = x_i$ if x_i is uncensored and $w_i = y_i = u_i$ if x_i is (right) censored at u_i. Let the deductible be d_i where $d_i = 0$ means no (left) truncation.

obs	w_i	d_i	$L(\boldsymbol{\theta}\|w_i)$	comment
1	$y = 0.1$	0	$1 - F(0.1)$	censored at 0.1
2	$x = 0.8$	0	f(0.8)	uncensored
3	$y = 5$	0.3	$\dfrac{1 - F(5)}{1 - F(0.3)}$	censored at 5 and truncated at 0.3
4	$x = 4.1$	1	$\dfrac{f(4.1)}{1 - F(1)}$	uncensored but truncated at 1

For the above table, $L(\boldsymbol{\theta}) = [1 - F(0.1)][f(0.8)] \left[\dfrac{1 - F(5)}{1 - F(0.3)}\right] \left[\dfrac{f(4.1)}{1 - F(1)}\right]$.

49) When finding $L(\boldsymbol{\theta}) = L(\boldsymbol{\theta}|x_1, ..., x_n) = \prod_{i=1}^n L(\boldsymbol{\theta}|x_i)$, consider the numerator and denominator for the term $L(\boldsymbol{\theta}|x_i)$. a) For the numerator, use $f(x_i)$ if x_i is known. If it is only known that x_i is between y and z use $F(z) - F(y)$ (where X is continuous with grouped data or right censored at u with $y = u$ and $z = \infty$ so $F(z) - F(y) = 1 - F(u)$). b) Let d be the truncation point. Then the denominator is $1 - F(d)$. (Use $d = 0$ if there is no truncation so $1 - F(0) = 1$.)

50) For individual discrete data, $L(\boldsymbol{\theta}) = \prod_{i=1}^n p(x_i) = \prod_{j=1}^m [p(x_j)]^{n_j}$ where n_j is the number of times $x_i = x_j$ for $j = 1, ..., m$. This is useful if the discrete data is tabled for m values of x_j.

51) For independent data $X_1, ..., X_n$, $L(\boldsymbol{\theta}) = \prod_{i=1}^n f_{X_i}(x_i)$ or $L(\boldsymbol{\theta}) = \prod_{i=1}^n p_{X_i}(x_i)$ where the X_i need not have the same distribution. Then $\ln[L(\boldsymbol{\theta})] = \sum_{i=1}^n \ln[f_{X_i}(x_i)]$ or

$\ln[L(\boldsymbol{\theta})] = \sum_{i=1}^n \ln[p_{X_i}(x_i)]$.

52) If there is a *deductible* d, and a *policy limit = maximum payment* $u - d$, then the *maximum covered loss* $u = u - d + d$. For a *coinsurance* policy, the insurance company pays αX of the loss for some $\alpha \in (0, 1]$. For coinsurance with a deductible, the insurance company pays $\alpha(X - d)_+$.

53) Using the above notation if Y^P is the per payment RV and X is the loss RV, then $Y^P = \alpha(X - d)$, so $X = \frac{Y^P}{\alpha} + d$. Here X is (left) truncated at d and (right) censored at u. Note $d = 0$, $u = \infty$, and $\alpha = 1$ are possible (no deductible, no maximum payment or coverage, no coinsurance). Note that Y^P is (right) censored at $\alpha(u - d)$.

54) Given losses X_i, but told that there is a deductible d and maximum payment $u - d$ or maximum covered loss $u = u - d + d$, be able to convert the loss data for X_i to what you would have based on insurer payments $Y^P = \min((X - d)_+, u)$ provided $Y^P > 0$. Then convert the data into losses the insurer would observe using 53) to find the MLE. See HW4 3b).

55) Let $f(x) = f(x|\boldsymbol{\theta})$ be the pdf or pmf of X. Assume the support of X does not depend on any unknown parameters. (See 37).) The family of pdfs or pmfs is a k parameter exponential family if

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

and k is the smallest integer where (1) holds.

56) For a one parameter exponential family, $f(x|\theta) = h(x)c(\theta) \exp[w(\theta)t(x)]$. Let $\eta = w(\theta)$ and let Ω be the parameter space (range) of η. If Ω is an open interval (a, b) (with $a = -\infty$ and $b = \infty$ possible), then X is from a one parameter regular exponential family (1PREF).

57) For a two parameter exponential family, $f(x|\boldsymbol{\theta}) = h(x)c(\boldsymbol{\theta})\exp\left[w_1(\boldsymbol{\theta})t_1(x) + w_2(\boldsymbol{\theta})t_2(x)\right]$. Let $\eta_1 = w_1(\boldsymbol{\theta})$ and $\eta_2 = w_2(\boldsymbol{\theta})$. If i) the parameter space of $(\eta_1, \eta_2) = \Omega$ is a cross product of two open intervals, and ii) neither η_1 and η_2 nor $t_1(x)$ and $t_2(x)$ satisfy a linearity constraint, then X is from a two parameter regular exponential family (2PREF).

58) Suppose X is from a 2PREF. If one of the two parameters is known, then X is from a 1PREF.

59) All of the brand name exponential families are regular except the inverse Gaussian distribution. 1PREFs: EXP(θ), Poisson(λ), binomial(q, m) with m known, single parameter Pareto(α, θ) with θ known, Weibull(θ, τ) with τ known, inverse exponential (θ), two parameter Pareto(α, θ) with θ known, inverse Weibull(θ, τ) with τ known, geometric(β), negative binomial(β, r) with r known. 2PREFs: $N(\mu, \sigma^2)$, $LN(\mu, \sigma)$, Gamma(α, θ), beta(a, b).

The $U(0, \theta)$ distribution is not an exponential family since the support depends on θ.

60) Let $\boldsymbol{x} = (x_1, ..., x_n)$. Let $\boldsymbol{X} = (X_1, ..., X_n)$. Let $\boldsymbol{X}$ have joint pdf or pmf $f(\boldsymbol{x}|\theta) = L(\theta)$. As a RV, let the likelihood $L(\theta) = L(\theta|\boldsymbol{X})$. Then Fisher's information or the information number or the information is $I_n(\theta) = I(\theta) = E_\theta\left(\left[\frac{d}{d\theta}\ln(L(\theta))\right]^2\right) = E_\theta\left(\left[\frac{d}{d\theta}\ln(f(\boldsymbol{X}|\theta))\right]^2\right)$. Then $I_1(\theta) = E_\theta\left(\left[\frac{d}{d\theta}\ln(L(\theta|X))\right]^2\right) = E_\theta\left(\left[\frac{d}{d\theta}\ln(f(X|\theta))\right]^2\right)$.

61) Let $\boldsymbol{\theta} = (\theta_1, ..., \theta_k)$. Then the $k \times k$ information matrix $\boldsymbol{I}_n(\boldsymbol{\theta}) = \boldsymbol{I}(\boldsymbol{\theta}) = (I_{ij})$ where $I_{ij} = E\left[\frac{\partial}{\partial\theta_i}\ln(L(\boldsymbol{\theta}))\frac{\partial}{\partial\theta_j}\ln(L(\boldsymbol{\theta}))\right]$ where $L(\boldsymbol{\theta}) = L(\boldsymbol{\theta}|\boldsymbol{X})$. If there are $k = 2$ parameters,

$$\boldsymbol{I}_n(\boldsymbol{\theta}) = \boldsymbol{I}(\boldsymbol{\theta}) = \begin{bmatrix} E\left(\left[\frac{\partial}{\partial\theta_1}\ln(L(\boldsymbol{\theta}))\right]^2\right) & E\left[\frac{\partial}{\partial\theta_1}\ln(L(\boldsymbol{\theta}))\frac{\partial}{\partial\theta_2}\ln(L(\boldsymbol{\theta}))\right] \\ E\left[\frac{\partial}{\partial\theta_2}\ln(L(\boldsymbol{\theta}))\frac{\partial}{\partial\theta_1}\ln(L(\boldsymbol{\theta}))\right] & E\left(\left[\frac{\partial}{\partial\theta_2}\ln(L(\boldsymbol{\theta}))\right]^2\right) \end{bmatrix}.$$

Since $\boldsymbol{I}(\boldsymbol{\theta})$ is symmetric, only 3 terms need to be computed when $k = 2$.

62) If $X_1, ..., X_n$ are iid from a k–parameter exponential family, then $I_{ij} = -E\left[\frac{\partial^2}{\partial\theta_i\partial\theta_j}\ln(L(\boldsymbol{\theta}))\right]$. If $k = 1$ then $I_n(\theta) = -E\left[\frac{d^2}{d\theta^2}\ln(L(\theta|\boldsymbol{X}))\right] = nI_1(\theta)$ where $I_1(\theta) = -E\left[\frac{d^2}{d\theta^2}\ln(L(\theta|X))\right]$ where X is a RV since $n = 1$. Hence $I_1(\theta) = -E\left[\frac{d^2}{d\theta^2}\ln(f(X|\theta))\right]$ where $f(x|\theta)$ is the pdf or pmf of X. If $k = 2$, then

$$\boldsymbol{I}_n(\boldsymbol{\theta}) = \boldsymbol{I}(\boldsymbol{\theta}) = -\begin{bmatrix} E\left[\frac{\partial^2}{\partial\theta_1^2}\ln(L(\boldsymbol{\theta}))\right] & E\left[\frac{\partial^2}{\partial\theta_1\partial\theta_2}\ln(L(\boldsymbol{\theta}))\right] \\ E\left[\frac{\partial^2}{\partial\theta_2\partial\theta_1}\ln(L(\boldsymbol{\theta}))\right] & E\left[\frac{\partial^2}{\partial\theta_2^2}\ln(L(\boldsymbol{\theta}))\right] \end{bmatrix}.$$

Since $\boldsymbol{I}(\boldsymbol{\theta})$ is symmetric, only 3 terms need to be computed when $k = 2$.

63) If $X_1, ..., X_n$ are iid, then under regularity conditions, if $k = 1$ then $I_1(\theta) = -E\left[\frac{d^2}{d\theta^2}\ln(f(X|\theta))\right]$, if $k = 2$ then

$$\boldsymbol{I}_1(\boldsymbol{\theta}) = -\begin{bmatrix} E\left[\frac{\partial^2}{\partial\theta_1^2}\ln(f(X|\boldsymbol{\theta}))\right] & E\left[\frac{\partial^2}{\partial\theta_1\partial\theta_2}\ln(f(X|\boldsymbol{\theta}))\right] \\ E\left[\frac{\partial^2}{\partial\theta_2\partial\theta_1}\ln(f(X|\boldsymbol{\theta}))\right] & E\left[\frac{\partial^2}{\partial\theta_2^2}\ln(f(X|\boldsymbol{\theta}))\right] \end{bmatrix},$$

and $I_n(\theta) = nI_1(\theta)$ or $\boldsymbol{I}_n(\boldsymbol{\theta}) = n\boldsymbol{I}_1(\boldsymbol{\theta})$.

64) **Unless told otherwise**, if $X_1, ..., X_n$ are iid, assume equations 62) and 63) can be used. **Exception**: 60) must be used instead of 62) if $X_1, ..., X_n$ are iid $U(0, \theta)$.

65) A $k \times 1$ random vector $\boldsymbol{X}$ has a $k-$dimensional *multivariate normal distribution* $\boldsymbol{X} \sim N_k(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ iff $\boldsymbol{t}^T \boldsymbol{X}$ has a univariate normal distribution for any $k \times 1$ constant vector $\boldsymbol{t}$. If $\boldsymbol{\Sigma}$ is positive definite, then $\boldsymbol{X}$ has a joint pdf

$$f(\boldsymbol{z}) = \frac{1}{(2\pi)^{k/2}|\boldsymbol{\Sigma}|^{1/2}} e^{-(1/2)(\boldsymbol{z}-\boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\boldsymbol{z}-\boldsymbol{\mu})}$$

where $|\boldsymbol{\Sigma}|^{1/2}$ is the square root of the determinant of $\boldsymbol{\Sigma}$.

66) If $\boldsymbol{X} \sim N_k(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, then $E(\boldsymbol{X}) = \boldsymbol{\mu}$ and $Cov(\boldsymbol{X}) = \boldsymbol{\Sigma}$ where $E(\boldsymbol{X}) = (E(X_1), ..., E(X_k))^T$ and $Cov(\boldsymbol{X}) = \boldsymbol{\Sigma} = (\sigma_{ij})$ where $\sigma_{ij} = Cov(X_i, X_j)$. Note that $\sigma_{ii} = \sigma_i^2 = V(X_i)$.

67) **MLE Limit Theorem:** Let $X_1, ..., X_n$ be iid. a) Let $\hat{\theta}_n$ be the MLE of θ. Under regularity conditions and if $\tau'(\theta) \neq 0$, then

$$\sqrt{n}[\hat{\theta}_n - \theta] \xrightarrow{D} N\left(0, \frac{1}{I_1(\theta)}\right) \quad \text{and} \quad \sqrt{\mathrm{n}}[\tau(\hat{\theta}_\mathrm{n}) - \tau(\theta)] \xrightarrow{\mathrm{D}} \mathrm{N}\left(0, \frac{[\tau'(\theta)]^2}{\mathrm{I}_1(\theta)}\right).$$

So $\hat{\theta} \sim AN\left(\theta, \frac{1}{nI_1(\theta)}\right)$ and $\tau(\hat{\theta}) \sim AN\left(\tau(\theta), \frac{[\tau'(\theta)]^2}{nI_1(\theta)}\right)$.

b) Let $\hat{\boldsymbol{\theta}}_n$ be the MLE of $\boldsymbol{\theta}$. Then under regularity conditions,

$$\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}) \xrightarrow{D} N_k\left(0, \boldsymbol{I}_1^{-1}(\boldsymbol{\theta})\right) \quad \text{and} \quad \hat{\boldsymbol{\theta}} \sim \mathrm{AN_k}\left(\boldsymbol{\theta}, \frac{\boldsymbol{I}_1^{-1}(\boldsymbol{\theta})}{\mathrm{n}}\right) = \mathrm{AN_k}(\boldsymbol{\theta}, \boldsymbol{I}_\mathrm{n}^{-1}(\boldsymbol{\theta})).$$

68) If $\boldsymbol{A} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$, then $\boldsymbol{A}^{-1} = \frac{1}{a_{11}a_{22} - a_{21}a_{12}} \begin{bmatrix} a_{22} & -a_{12} \\ -a_{21} & a_{11} \end{bmatrix}$. In this class $\boldsymbol{A}$ is usually symmetric ($\boldsymbol{A} = \boldsymbol{A}^T$), so $a_{12} = a_{21}$.

69) **Delta Method**: If $g'(\theta) \neq 0$ and $\sqrt{n}(T_n - \theta) \xrightarrow{D} N(0, \sigma^2)$, then $\sqrt{n}(g(T_n) - g(\theta)) \xrightarrow{D} N(0, \sigma^2[g'(\theta)]^2)$. For example, apply to the CLT with $T_n = \overline{X}$ and $\theta = \mu$ or MLE limit theorem 67a) with $T_n = \hat{\theta}_n$ and $g(\theta) = \tau(\theta)$.

70) The asymptotic variance of the MLE $\hat{\theta}$ is $Var(\hat{\theta}) = \frac{1}{nI_1(\theta)} = \frac{1}{I_n(\theta)}$. The asymptotic variance of $\tau(\hat{\theta})$ is $Var(\tau(\hat{\theta})) = \frac{[\tau'(\theta)]^2}{nI_1(\theta)} = \frac{[\tau'(\theta)]^2}{I_n(\theta)}$ if $\tau'(\theta) \neq 0$. The asymptotic covariance matrix of the MLE $\hat{\boldsymbol{\theta}}$ is $Cov(\hat{\boldsymbol{\theta}}) = \boldsymbol{I}_n^{-1}(\boldsymbol{\theta}) = [n\boldsymbol{I}_1(\boldsymbol{\theta})]^{-1}$ when $\hat{\theta}$ or $\hat{\boldsymbol{\theta}}$ is asymptotically normal as in 67). The asymptotic variance of $\hat{\theta}$ approximates the variance of $\hat{\theta}$ and the asymptotic covariance matrix of $\hat{\boldsymbol{\theta}}$ approximates the covariance matrix of $\hat{\boldsymbol{\theta}}$. If $g'(\theta) \neq 0$ and $Var(\hat{\theta})$ is the asymptotic variance of $\hat{\theta}$, then the the Delta Method asymptotic variance of $g(\hat{\theta})$ is $Var(g(\hat{\theta})) = [g'(\theta)]^2 Var(\hat{\theta})$.

71) Suppose $X_1, ..., X_n$ are iid with mean $E(X)$ and variance $V(X)$ where $\theta = cE(X)$ and the MLE $\hat{\theta} = c\overline{X}$ for some known constant c. Then $I_1(\theta) = \frac{1}{c^2 V(X)}$ and the asymptotic variance $Var(\hat{\theta}) = V(c\overline{X}) = c^2 V(X)/n$.

72) Let $g : \mathbb{R}^2 \to \mathbb{R}$ have nonzero partial derivatives. Then the Delta Method asymptotic variance of $g(\hat{\theta}_1, \hat{\theta}_2)$ is $Var(g(\hat{\theta}_1, \hat{\theta}_2)) =$

$$\left(\frac{\partial g(\boldsymbol{\theta})}{\partial \theta_1}\right)^2 Var(\hat{\theta}_1) + 2\left(\frac{\partial g(\boldsymbol{\theta})}{\partial \theta_1}\right)\left(\frac{\partial g(\boldsymbol{\theta})}{\partial \theta_2}\right) Cov(\hat{\theta}_1, \hat{\theta}_2) + \left(\frac{\partial g(\boldsymbol{\theta})}{\partial \theta_2}\right)^2 Var(\hat{\theta}_2)$$

where the asymptotic covariance matrix of $(\hat{\theta}_1, \hat{\theta}_2)$ is $Cov(\hat{\boldsymbol{\theta}}) = \begin{bmatrix} Var(\hat{\theta}_1) & Cov(\hat{\theta}_1, \hat{\theta}_2) \\ Cov(\hat{\theta}_1, \hat{\theta}_2) & Var(\hat{\theta}_2) \end{bmatrix}$, where $Cov(\hat{\boldsymbol{\theta}}) = \boldsymbol{I}_n^{-1}(\boldsymbol{\theta})$ if $\hat{\boldsymbol{\theta}}$ is the MLE. It is also true that

$$Var(g(\hat{\theta}_1, \hat{\theta}_2)) = \left(\frac{\partial g(\boldsymbol{\theta})}{\partial \theta_1} \quad \frac{\partial g(\boldsymbol{\theta})}{\partial \theta_2} \right) \begin{bmatrix} Var(\hat{\theta}_1) & Cov(\hat{\theta}_1, \hat{\theta}_2) \\ Cov(\hat{\theta}_1, \hat{\theta}_2) & Var(\hat{\theta}_2) \end{bmatrix} \begin{pmatrix} \frac{\partial g(\boldsymbol{\theta})}{\partial \theta_1} \\ \\ \frac{\partial g(\boldsymbol{\theta})}{\partial \theta_2} \end{pmatrix}.$$

Mnemonic: $Cov(\hat{\theta}_i, \hat{\theta}_j)$ has coefficient $\frac{\partial g(\boldsymbol{\theta})}{\partial \theta_i} \ \frac{\partial g(\boldsymbol{\theta})}{\partial \theta_j}$ where $Cov(\hat{\theta}_i, \hat{\theta}_i) = Var(\hat{\theta}_i)$.

73) Let $X_1, ..., X_n$ be iid with distribution X. Here are some values of $I_1(\theta)$ and the asymptotic variance $Var(\hat{\theta})$ of the MLE. See 45).

a) If $X \sim EXP(\theta)$, then $\hat{\theta} = \overline{X}$, $I_1(\theta) = \frac{1}{\theta^2}$ and $Var(\hat{\theta}) = \frac{\theta^2}{n}$.

b) If $X \sim LN(\mu, \sigma)$ with μ and σ^2 are unknown. Let $W_i = \ln(X_i)$. Then $\hat{\mu} = \overline{W}$ and $\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (W_i - \overline{W})^2$. Now $I_1(\mu) = \frac{1}{\sigma^2}$, $I_1(\sigma) = \frac{2}{\sigma^2}$, $Var(\hat{\mu}) = \frac{\sigma^2}{n}$, $Var(\hat{\sigma}) = \frac{\sigma^2}{2n}$, and $Cov(\hat{\mu}, \hat{\sigma}) = 0$.

c) Let $X \sim$ Pareto(α, θ). i) If α is fixed, then $I_1(\theta) = \frac{\alpha}{(\alpha+2)\theta^2}$ and $Var(\hat{\theta}) = \frac{(\alpha+2)\theta^2}{n\alpha}$. ii) If θ is fixed, then $I_1(\alpha) = \frac{1}{\alpha^2}$ and $Var(\hat{\alpha}) = \frac{\alpha^2}{n}$.

d) If $X \sim$ Weibull(θ, τ) with τ fixed, then $I_1(\theta) = \frac{\tau^2}{\theta^2}$ and $Var(\hat{\theta}) = \frac{\theta^2}{n\tau^2}$.

e) If $X \sim U(0, \theta)$ then $\hat{\theta} = X_{(n)}$, $I_1(\theta) = \frac{1}{\theta^2}$ and $Var(\hat{\theta}) = \frac{n\theta^2}{(n+1)^2(n+2)} \neq \frac{1}{nI_1(\theta)}$, and $\hat{\theta}$ is not asymptotically normal.

f) If $X \sim N(\mu, \sigma^2)$ with μ and σ^2 unknown, then $\hat{\mu} = \overline{X}$, $\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \overline{X})^2$, $I_1(\mu) = \frac{1}{\sigma^2}$, $I_1(\sigma^2) = \frac{1}{2\sigma^4}$, $Var(\hat{\mu}) = \frac{\sigma^2}{n}$, $Var(\hat{\sigma}^2) = \frac{2\sigma^4}{n}$, and $Cov(\hat{\mu}, \hat{\sigma}^2) = 0$.

g) If $X \sim$ Poisson(λ) then $\hat{\lambda} = \overline{X}$, $I_1(\lambda) = \frac{1}{\lambda}$ and $Var(\hat{\lambda}) = \frac{\lambda}{n}$.

h) If $X \sim$ binomial(q, k), k **known**, then $\hat{q} = \overline{X}/k$, $I_1(q) = \frac{k}{q(1-q)}$ and $Var(\hat{q}) = \frac{q(1-q)}{nk}$.

i) If $X \sim$ negative binomial (r, β) with r **known**, then $\hat{\beta} = \frac{\overline{X}}{r}$, $I_1(\beta) = \frac{r}{\beta(1+\beta)}$ and $Var(\hat{\beta}) = \frac{\beta(1+\beta)}{nr}$.

j) If $X \sim G(\alpha, \theta)$ with α **known**, then $\hat{\theta} = \overline{X}/\alpha$, $I_1(\theta) = \frac{\alpha}{\theta^2}$ and $Var(\hat{\theta}) = \frac{\theta^2}{n\alpha}$.

k) If $X \sim$ geometric(β), then $\hat{\beta} = \overline{X}$, $I_1(\beta) = \frac{1}{\beta(1+\beta)}$ and $Var(\hat{\beta}) = \frac{\beta(1+\beta)}{n}$.

74) **Know:** When asked to find $Var(\hat{\theta})$, $Var(g(\hat{\theta}_1, \hat{\theta}_2))$, and $Cov(\hat{\boldsymbol{\theta}})$, plug in $\hat{\theta}, \hat{\theta}_1, \hat{\theta}_2$, or $\hat{\boldsymbol{\theta}}$ to get real numbers or a matrix with real number entries.

75) If

$$\sqrt{n} \begin{pmatrix} \hat{\theta}_1 - \theta_1 \\ \hat{\theta}_2 - \theta_2 \end{pmatrix} \xrightarrow{D} N_2 \left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{pmatrix} \right],$$

then $\sqrt{n}(\hat{\theta}_i - \theta_i) \xrightarrow{D} N(0, \sigma_{ii})$ for $i = 1, 2$ where the asymptotic variance $Var(\hat{\theta}_i) = \sigma_{ii}/n$.

76) Let z_p be the $1 - \alpha/2$ percentile $z_{1-\alpha/2}$ = the upper $\alpha/2$ percentile $z_{\alpha/2}$, using bad notation. So $P(Z \leq z_p) = 1 - \alpha/2$ and $P(Z > z_p) = \alpha/2$.

CI	90%	95%	99%
z_p	1.645	1.96	2.576

77) Using the notation in 76), if $\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{D} N(0, \sigma^2)$, then $Var(\hat{\theta}) = \sigma^2/n$, and a large sample $100(1 - p/2)\%$ confidence interval (CI) for θ is

$$\hat{\theta} \pm z_p \sqrt{\widehat{Var}(\hat{\theta})} = (\hat{\theta} - z_p \frac{\hat{\sigma}}{\sqrt{n}}, \hat{\theta} + z_p \frac{\hat{\sigma}}{\sqrt{n}}).$$

If $\hat{\theta}$ is the MLE, then the $100(1 - p/2)\%$ CI for θ is

$$\hat{\theta} \pm z_p \frac{1}{\sqrt{n I_1(\hat{\theta})}} = \hat{\theta} \pm z_p \frac{1}{\sqrt{I_n(\hat{\theta})}}.$$

The $100(1 - p/2)\%$ CI for $\tau(\theta) = g(\theta)$ is

$$\tau(\hat{\theta}) \pm z_p \sqrt{\frac{[\tau'(\hat{\theta})]^2}{n I_1(\hat{\theta})}} = \tau(\hat{\theta}) \pm z_p \sqrt{\frac{[\tau'(\hat{\theta})]^2}{I_n(\hat{\theta})}} = \tau(\hat{\theta}) \pm z_p \sqrt{\widehat{Var}(\tau(\hat{\theta}))}.$$

The $100(1 - p/2)\%$ CI for $g(\theta_1, \theta_2)$ is

$$g(\hat{\theta}_1, \hat{\theta}_2) \pm z_p \sqrt{\widehat{Var}(g(\hat{\theta}_1, \hat{\theta}_2))}$$

where $\widehat{Var}(g(\hat{\theta}_1, \hat{\theta}_2))$ plugs $(\hat{\theta}_1, \hat{\theta}_2)$ in for unknowns (θ_1, θ_2) in the $Var(g(\hat{\theta}_1, \hat{\theta}_2))$ formula.

78) Let m = number of uncensored observations, c = number of censored observations, $n = m + c$, let d_i be the truncation point for each observation (0 if untruncated). Let x_i be the observation if uncensored or the censoring point (u_i) if censored. The following formulas work if left truncation and right censoring are present or not. For a), see 53).

a) EXP(θ): $\hat{\theta} = \frac{\sum_{i=1}^n (x_i - d_i)}{m} = \frac{\sum_{i=1}^n Y_i^P}{m}$.

b) Weibull fixed τ: $\hat{\theta} = \left(\frac{\sum_{i=1}^n (x_i^\tau - d_i^\tau)}{m} \right)^{1/\tau}$.

c) Pareto fixed θ: $\hat{\alpha} = \frac{-m}{\sum_{i=1}^n \ln \left(\frac{\theta + d_i}{\theta + x_i} \right)}$.

d) single parameter Pareto fixed θ: $\hat{\alpha} = \frac{-m}{\sum_{i=1}^n \ln \left(\frac{\max(\theta, d_i)}{x_i} \right)}$.

79) Suppose the $U(0,\theta)$ distribution is used for grouped or censored data. Let $[c, f)$ be the highest integer with count $n_c > 0$ ($f = \infty$ is allowed and c could be the censoring value). Let m = number of observations $< c$, so $n = m + n_c$. Then $\hat{\theta} = \min\left(\frac{n}{m} c, f\right)$. Note that $\hat{\theta}$ is found by matching $P(X < c) = p \stackrel{set}{=} \hat{p} = \frac{m}{n}$. For censored data use the table below.

interval	number
$(0, c)$	m
$[c, \infty)$	n_c

For grouped data with $c = c_{k-1}$, $n = \sum_{i=1}^k n_i$, and $m = n - n_k$, use either of the two tables below.

interval	number	or	interval	number
(c_0, c_1)	n_1		(c_0, c_1)	n_1
(c_1, c_2)	n_2		(c_1, c_2)	n_2
(c_2, c_3)	n_3		(c_2, c_3)	n_3
$\vdots$	$\vdots$		$\vdots$	$\vdots$
(c_{k-1}, c_k)	$n_k > 0$		(c_{k-1}, c_k)	$n_k > 0$
			(c_k, ∞)	0

80) The **Bernoulli technique** is useful if there are 2 classes or groups, and $p = p(\theta) = P(X$ is in the 1st group) has $\theta = p^{-1}(p(\theta))$. Then $L(p) = p^{n_1}(1-p)^{n_2}$, $\hat{p} = n_1/n$, $1 - \hat{p} = n_2/n$. Solve $p \stackrel{set}{=} \hat{p} = n_1/n$ for θ to get the MLE $\hat{\theta}$. For a discrete distribution, if "0 claims" is the first class, then $p = p_0 = P(X = 0)$. If X is continuous, then $p = F(c)$ and $1-p = S(c)$ for the table below. Two typical tables are below. Note that $n = n_1 + n_2$.

class	or	class	number
0 claims		$(0, c)$	n_1
1 or more claims		(c, ∞)	n_2

81) The Poisson, binomial, negative binomial and Geometric distributions are the only members of the $(a, b, 0)$ class. X is a member of this class if $\frac{p_k}{p_{k-1}} = a + \frac{b}{k}$ for $k = 1, 2, ...$ Hence $\frac{k\ p_k}{p_{k-1}} = a\ k + b$ for $k = 1, 2, ...$ except the recursion goes up to $k = m$ for the binomial. In a sample (e.g. 0,1,1,5,0,3,7,0,5,2,1,1,1,4,2,2, or see 82) or 83)), let n_k = number in sample equal to k. Plot k versus $\frac{k\ \hat{p}_k}{\hat{p}_{k-1}} = \frac{k\ n_k}{n_{k-1}}$ where k is omitted if $n_k = 0$ (or if either count n_k or n_{k-1} is small). If the n_k are large, the plot should follow a straight line with slope a. Often only need to compute the first 3 or 4 terms.

The Poisson RV X has slope $a = 0$ and $E(X) = V(X)$. The bin RV X has slope $a < 0$ and $E(X) > V(X)$. The NB RV X (Geometric is NB($\beta, r = 1$)) has slope $a > 0$ and $E(X) < V(X)$. Often N is used instead of X.

For the Poisson RV, won't get $\overline{X} = \hat{\sigma}^2_E$ or $\overline{X} = \hat{\sigma}^2_U$, but $\dfrac{k\ n_k}{n_{k-1}}$ are roughly constant in that they oscillate about a number (rather than clearly increase or decrease) if the n_k and n_{k-1} are large. If choosing from the $(a, b, 0)$ class and $\overline{X}$ is clearly larger than $\hat{\sigma}^2$, choose the binomial distribution. If $\overline{X}$ is clearly smaller than $\hat{\sigma}^2$, then choose a negative binomial distribution, possibly the Geometric distribution. (If a 90% CI for E(X) contains $\hat{\sigma}^2$, then we might say that it is not clear that $\overline{X}$ is larger or smaller than $\hat{\sigma}^2$.) Note that $p_0 = P(X = 0)$ and $p_k = P(X = k)$.

dist	a	b	p_0	
Poisson(λ)	0	λ	$e^{-\lambda}$	$E(X) = V(X)$
bin(q, m)	$\dfrac{-q}{1-q}$	$(m+1)\dfrac{q}{1-q}$	$(1-q)^m$	$E(X) > V(X)$
NB(β, r)	$\dfrac{\beta}{1+\beta}$	$(r-1)\dfrac{\beta}{1+\beta}$	$(1+\beta)^{-r}$	$E(X) < V(X)$
Geom(β)	$\dfrac{\beta}{1+\beta}$	0	$(1+\beta)^{-1}$	$E(X) < V(X)$

82) For complete tabled discrete data, $L = \prod_{k=0}^m p_k^{n_k}$, $\ln(L) = \sum_{k=0}^m n_k \ln(p_k)$, $\dfrac{1}{n}\sum_{i=1}^n X_i^j = \sum_{i=1}^m k^j n_k$ where often $j = 1$ or $j = 2$. Recall that $\overline{X}$ uses $j = 1$, the biased variance or empirical variance $\hat{\sigma}^2_E = \dfrac{1}{n}\sum_{i=1}^n X_i^2 - (\overline{X})^2$ and unbiased (sample) variance $\hat{\sigma}^2_U = \dfrac{n}{n-1}\hat{\sigma}^2_E$.

k	n_k	or	n_k
0	n_0		n_0
1	n_1		n_1
2	n_2		n_2
$\vdots$	$\vdots$		$\vdots$
m	n_m		n_m
$(m+1)+$	0		$n^+_{m+1} > 0$

83) For tabled discrete data, where the last class has a count $n^+_{m+1} > 0$ equal to the number of observations $\geq m+1$, $L = (\prod_{k=0}^m p_k^{n_k})(1 - p_0 - \cdots - p_m)^{n^+_{m+1}}$.

ch 16

84) Let X have cdf F and pdf f. Let f_n be the empirical pdf and let the empirical cdf $F_n(x) = \dfrac{1}{n}\sum_{i=1}^n I(X_i \leq x)$ where the indicator function $I(X_i \leq x) = 1$ if $X_i \leq x$ and $I(X_i \leq x) = 0$ if $X_i > x$. Then the indicator variables are iid binomial($q = F(x), m = 1$). If X is left truncated at d, let $F^*(x) = \dfrac{F(x) - F(d)}{1 - F(d)}$ and $f^*(x) = \dfrac{f(x)}{1 - F(d)}$ be the cdf and pdf of the truncated RV for $x \geq d$.

85) Our convention will be that a plot of a versus b will have a on the horizontal axis and b on the vertical axis.

86) Suppose $X_1, ..., X_n$ are iid. A $D(x)$ plot is a plot of x versus $D(x) = F_n(x) - F^*(x)$ where $F^*(x)$ is often found using the MLE. Want the plot to oscillate about the $D(x) = 0$ line (often the horizontal axis).

87) Suppose $X_1, ..., X_n$ are iid. Order the observations $x_1 \leq \cdots \leq x_n$ (actually the order statistics but x_i is used instead of $x_{(i)}$). The p–p plot is a plot of the sample CDF "$F_n(x_j) = \dfrac{j}{n+1}$" versus the fitted CDF $F^*(x_j)$ which is often found using the MLE. If n is large, then a good fit is indicated by the plotted points clustering tightly about the identity line with zero intercept and slope 1 that passes though the points (0,0) and (1,1).

88) Often the plotted points in the p–p plot are indicated by a curve (or approximate a curve). a) If the slope of the curve > 1 at a point, then in the region neighboring the point, the fitted distribution is "thick," or "heavy," or "has too much weight in the region," or "the fitted distribution has more probability in the region than the sample."
b) If the slope of the curve < 1 at a point, then in the region neighboring the point, the fitted distribution is "thin," or "light," or "has too little weight in the region," or "the fitted distribution has less probability in the region than the sample."
c) The "left tail" is the region near (0,0) (with $j/(n+1)$ near 0), and the "right tail" is the region near (1,1) (with $j/(n+1)$ near 1).
d) An "ess shape" suggests that the left and right tails of fitted distribution are too light and that the fitted distribution has too much probability in the middle of the plot where the "ess shaped" curve intersects the identity line.
e) Suppose the plot is the reflection of the plot in d) about the identity line. Such a plot indicates that the left and right tails of the fitted distribution are too heavy and that the fitted distribution has too little probability in the middle of the plot where the curve intersects the identity line.

89) H_0 is the null hypothesis while $H_A = H_1$ is the alternative hypothesis. Use $\alpha = 0.05$ if α is not given.

90) The 4 step χ^2 test of hypotheses is below.
i) H_0: the fitted distribution is good $\quad H_A$: not H_0
ii) test statistic $Q = \displaystyle\sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i}$
iii) The degrees of freedom $= k - r - 1$ where r is the number of estimated parameters (preferably MLEs), and $r = 0$ is possible. If $Q >$ cutoff, then reject H_0 (Q is in the critical region and pval $< \alpha$). If $Q <$ cutoff, then fail to reject H_0 (Q is not in the critical region and pval $> \alpha$).
iv) Give a nontechnical conclusion: "reject H_0" implies that the fitted distribution is not good, while "fail to reject H_0" implies that the fitted distribution is good (or that there is not enough evidence to conclude that the fitted distribution is not good).

91) A variant of the 4 step test is to find the smallest significance level at which H_0 is rejected. For example, find that "H_0 is rejected at the 0.005 significance level" or that "H_0 is rejected at the 5% (0.05) significance level but not at the 2.5% (0.025) significance

level.

92) The χ^2 table gives the cutoff = critical value. Find degrees of freedom $df = k - r - 1$. If Q is larger than any value on the df line, then reject H_0 at significance level 0.005. If Q is between two critical values in the df line, then reject H_0 at the significance level corresponding to the smaller critical value, but not at the significance level corresponding to the larger critical value. The significance level corresponds to $1 - P$ where P is given on the top line of the table. Hence if $df = 3$ and $Q = 8$, then $7.815 < Q < 9.348$. So reject H_0 at the 0.05 significance level but not at the 0.025 significance level.

93) Data for the χ^2 test is given in a table. Categories and counts $n_i = O_i$ are given. Sometimes probabilities p_i are given but sometimes need to be computed. The $E_i = np_i$ and χ^2 contributions $C_i = \dfrac{(O_i - E_i)^2}{E_i}$ need to be computed. Note that $n = \sum_{i=1}^k n_i = \sum_{i=1}^k O_i = \sum_{i=1}^k E_i$, $1 = \sum_{i=1}^k p_i$ and $Q = \sum_{i=1}^k \dfrac{(O_i - E_i)^2}{E_i} = \sum_{i=1}^k C_i = (\sum_{i=1}^k \dfrac{O_i^2}{E_i}) - n$. Often p_k and E_k are obtained by subtraction $p_k = 1 - p_1 - \cdots - p_{k-1}$ and $E_k = n - E_1 - \cdots - E_{k-1}$. If category i is the interval (c_{i-1}, c_i), then $p_i = F(c_i) - F(c_{i-1})$ where often F is fitted by MLE. If category i is the value j, then $p_i = P(X = j)$ which is often fitted by MLE. (Note that $P(X = j)$ is often denoted by p_j for brand name discrete distributions.)

category	$O_i = n_i$	p_i	$E_i = np_i$	$C_i = \frac{(O_i - E_i)^2}{E_i}$
1	$O_1 = n_1$	p_1	E_1	C_1
2	$O_2 = n_2$	p_2	E_2	C_2
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
k	$O_k = n_k$	p_k	E_k	C_k
sum	n	1	n	Q

94) Want $E_i \geq 5$ for $i = 1, ..., k$. If necessary, combine neighboring classes so this result holds, unless the problem says to use the given classes. The last class for continuous data should be (c, f) where $f = \infty$ unless the support of the distribution is $(0, f)$. For discrete data the last class is $j+$ unless the distribution only has support on 0, 1, ..., j.

95) The $df = k - r - 1$ occurs if the n claims are divided into k groups since then $\sum_{i=1}^k n_i = n$ (conditional on n or after gathering the data). (Occasionally the claims are just observed in k groups, so n is a random variable or the groups are independent. Then $df = k - r$.)