# Solutions to Problems

## Lecture 1

1. $P\{\max(X,Y,Z) \le t\} = P\{X \le t \quad \text{and} \quad Y \le t \quad \text{and} \quad Z \le t\} = P\{X \le t\}^3$ by independence. Thus the distribution function of the maximum is $(t^6)^3 = t^{18}$, and the density is $18t^{17}, 0 \le t \le 1$.
2. See Figure S1.1. We have

$$P\{Z \le z\} = \int\int_{y \le zx} f_{XY}(x,y)\,dx\,dy = \int_{x=0}^{\infty}\int_{y=0}^{zx} e^{-x}e^{-y}\,dy\,dx$$

$$F_Z(z) = \int_0^\infty e^{-x}(1-e^{-zx})\,dx = 1 - \frac{1}{1+z}, \quad z \ge 0$$

$$f_Z(z) = \frac{1}{(z+1)^2}, \quad z \ge 0$$

   $F_Z(z) = f_Z(z) = 0$ for $z < 0$.
3. $P\{Y = y\} = P\{g(X) = y\} = P\{X \in g^{-1}(y)\}$, which is the number of $x_i$'s that map to $y$, divided by $n$. In particular, if $g$ is one-to-one, then $p_Y(g(x_i)) = 1/n$ for $i = 1, \dots, n$.
4. Since the area under the density function must be 1, we have $ab^3/3 = 1$. Then (see Figure S1.2) $f_Y(y) = f_X(y^{1/3})/|dy/dx|$ with $y = x^3, dy/dx = 3x^2$. In $dy/dx$ we substitute $x = y^{1/3}$ to get

$$f_Y(y) = \frac{f_X(y^{1/3})}{3y^{2/3}} = \frac{3}{b^3}\frac{y^{2/3}}{3y^{2/3}} = \frac{1}{b^3}$$

   for $0 < y^{1/3} < b$, i.e., $0 < y < b^3$.
5. Let $Y = \tan X$ where $X$ is uniformly distributed between $-\pi/2$ and $\pi/2$. Then (see Figure S1.3)

$$f_Y(y) = \frac{f_X(\tan^{-1} y)}{|dy/dx|_{x=\tan^{-1} y}} = \frac{1/\pi}{\sec^2 x}$$

   with $x = \tan^{-1} y$, i.e., $y = \tan x$. But $\sec^2 x = 1 + \tan^2 x = 1 + y^2$, so $f_Y(y) = 1/[\pi(1+y^2)]$, the Cauchy density.

## Lecture 2

1. We have $y_1 = 2x_1, y_2 = x_2 - x_1$, so $x_1 = y_1/2, x_2 = (y_1/2) + y_2$, and

$$\frac{\partial(y_1,y_2)}{\partial(x_1,x_2)} = \begin{vmatrix} 2 & 0 \\ -1 & 1 \end{vmatrix} = 2.$$

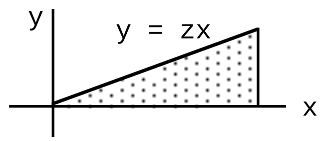


Figure S1.1

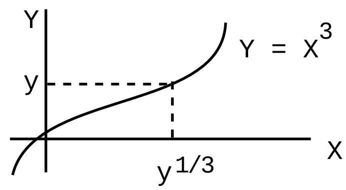


Figure S1.2

Thus $f_{Y_1Y_2}(y_1,y_2) = (1/2)f_{X_1X_2}(x_1,x_2) = e^{-x_1-x_2} = \exp[-(y_1/2)-(y_1/2)-y_2] = e^{-y_1}e^{-y_2}$. As indicated in the comments, the range of the $y$'s is $0 < y_1 < 1, 0 < y_2 < 1$. Therefore the joint density of $Y_1$ and $Y_2$ is the product of a function of $y_1$ alone and a function of $y_2$ alone, which forces independence.

2. We have $y_1 = x_1/x_2, y_2 = x_2$, so $x_1 = y_1y_2, x_2 = y_2$ and

$$\frac{\partial(x_1,x_2)}{\partial(y_1,y_2)} = \begin{vmatrix} y_2 & y_1 \\ 0 & 1 \end{vmatrix} = y_2.$$

Thus $f_{Y_1Y_2}(y_1,y_2) = f_{X_1X_2}(x_1,x_2)|\partial(x_1,x_2)/\partial(y_1,y_2)| = (8y_1y_2)(y_2)(y_2) = 2y_1(4y_2^3)$. Since $0 < x_1 < x_2 < 1$ is equivalent to $0 < y_1 < 1, 0 < y_2 < 1$, it follows just as in Problem 1 that $X_1$ and $X_2$ are independent.

3. The Jacobian $\partial(x_1,x_2,x_3)/\partial(y_1,y_2,y_3)$ is given by

$$\begin{vmatrix} y_2y_3 & y_1y_3 & y_1y_2 \\ -y_2y_3 & y_3-y_1y_3 & y_2-y_1y_2 \\ 0 & -y_3 & 1-y_2 \end{vmatrix}$$

$$= (y_2y_3^2 - y_1y_2y_3^2)(1-y_2) + y_1y_2^2y_3^2 + y_3(y_2-y_1y_2)y_2y_3 + (1-y_2)y_1y_2y_3^2$$

which cancels down to $y_2y_3^2$. Thus

$$f_{Y_1Y_2Y_3}(y_1,y_2,y_3) = \exp[-(x_1+x_2+x_3)]y_2y_3^2 = y_2y_3^2e^{-y_3}.$$

This can be expressed as $(1)(2y_2)(y_3^2e^{-y_3}/2)$, and since $x_1,x_2,x_3 > 0$ is equivalent to $0 < y_1 < 1, 0 < y_2 < 1, y_3 > 0$, it follows as before that $Y_1,Y_2,Y_3$ are independent.

## Lecture 3

1. $M_{X_2}(t) = M_Y(t)/M_{X_1}(t) = (1-2t)^{-r/2}/(1-2t)^{-r_1/2} = (1-2t)^{-(r-r_1)/2}$, which is $\chi^2(r-r_1)$.

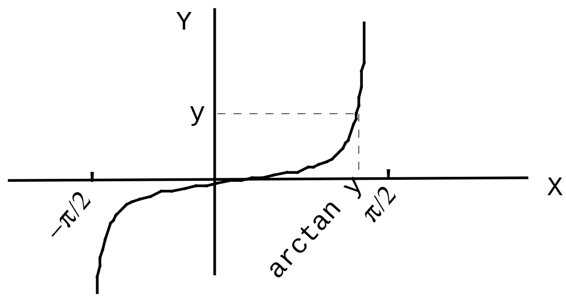


Figure S1.3

2. The moment-generating function of $c_1X_1 + c_2X_2$ is

$$E[e^{t(c_1X_1+c_2X_2)}] = E[e^{tc_1X_1}]E[e^{tc_2X_2}] = (1-\beta_1c_1t)^{-\alpha_1}(1-\beta_2c_2t)^{-\alpha_2}.$$

   If $\beta_1c_1 = \beta_2c_2$, then $X_1 + X_2$ is gamma with $\alpha = \alpha_1 + \alpha_2$ and $\beta = \beta_ic_i$.

3. $M(t) = E[\exp(\sum_{i=1}^n c_iX_i)] = \prod_{i=1}^n E[\exp(tc_iX_i)] = \prod_{i=1}^n M_i(c_it)$.
4. Apply Problem 3 with $c_i = 1$ for all $i$. Thus

$$M_Y(t) = \prod_{i=1}^n M_i(t) = \prod_{i=1}^n \exp[\lambda_i(e^t-1)] = \exp\left[\left(\sum_{i=1}^n \lambda_i\right)(e^t-1)\right]$$

   which is Poisson $(\lambda_1 + \cdots + \lambda_n)$.

5. Since the coin is unbiased, $X_2$ has the same distribution as the number of heads in the second experiment. Thus $X_1 + X_2$ has the same distribution as the number of heads in $n_1 + n_2$ tosses, namely binomial with $n = n_1 + n_2$ and $p = 1/2$.

## Lecture 4

1. Let $\Phi$ be the normal (0,1) distribution function, and recall that $\Phi(-x) = 1 - \Phi(x)$. Then

$$P\{\mu - c < \overline{X} < \mu + c\} = P\{-c\frac{\sqrt{n}}{\sigma} < \frac{\overline{X}-\mu}{\sigma/\sqrt{n}} < c\frac{\sqrt{n}}{\sigma}\}$$

$$= \Phi(c\sqrt{n}/\sigma) - \Phi(-c\sqrt{n}/\sigma) = 2\Phi(c\sqrt{n}/\sigma) - 1 \geq .954.$$

   Thus $\Phi(c\sqrt{n}/\sigma) \geq 1.954/2 = .977$. From tables, $c\sqrt{n}/\sigma \geq 2$, so $n \geq 4\sigma^2/c^2$.

2. If $Z = \overline{X} - \overline{Y}$, we want $P\{Z > 0\}$. But $Z$ is normal with mean $\mu = \mu_1 - \mu_2$ and variance $\sigma^2 = (\sigma_1^2/n_1) + (\sigma_2^2/n_2)$. Thus

$$P\{Z > 0\} = P\{\frac{Z-\mu}{\sigma} > \frac{-\mu}{\sigma}\} = 1 - \Phi(-\mu/\sigma) = \Phi(\mu/\sigma).$$

3. Since $nS^2/\sigma^2$ is $\chi^2(n-1)$, we have

$$P\{a < S^2 < b\} = P\{\frac{na}{\sigma^2} < \chi^2(n-1) < \frac{nb}{\sigma^2}\}.$$

If $F$ is the $\chi^2(n-1)$ distribution function, the desired probability is $F(nb/\sigma^2) - F(na/\sigma^2)$, which can be found using chi-square tables.

4. The moment-generating function is

$$E[e^{tS^2}] = E\left(\exp\left[\frac{nS^2}{\sigma^2}\frac{t\sigma^2}{n}\right]\right) = E[\exp(t\sigma^2 X/n)]$$

where the random variable $X$ is $\chi^2(n-1)$, and therefore has moment-generating function $M(t) = (1-2t)^{-(n-1)/2}$. Replacing $t$ by $t\sigma^2/n$ we get

$$M_{S^2}(t) = \left(1 - \frac{2t\sigma^2}{n}\right)^{-(n-1)/2}$$

so $S^2$ is gamma with $\alpha = (n-1)/2$ and $\beta = 2\sigma^2/n$.

## Lecture 5

1. By definition of the beta density,

$$E(X) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \int_0^\infty x^a (1-x)^{b-1}\, dx$$

and the integral is $\beta(a+1,b) = \Gamma(a+1)\Gamma(b)/\Gamma(a+b+1)$. Thus $E(X) = a/(a+b)$. Now

$$E(X^2) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \int_0^\infty x^{a+1} (1-x)^{b-1}\, dx$$

and the integral is $\beta(a+2,b) = \Gamma(a+2)\Gamma(b)/\Gamma(a+b+2)$. Thus

$$E(X^2) = \frac{(a+1)a}{(a+b+1)(a+b)}.$$

and

$$\operatorname{Var} X = E(X^2) - [E(X)]^2$$

$$= \frac{1}{(a+b)^2(a+b+1)}[(a+1)a(a+b) - a^2(a+b+1)] = \frac{ab}{(a+b)^2(a+b+1)}.$$

2. $P\{-c \le T \le c\} = F_T(c) - F_T(-c) = F_T(c) - (1 - F_T(c)) = 2F_T(c) - 1 = .95$, so $F_T(c) = 1.95/2 = .975$. From the $T$ table, $c = 2.131$.
3. $W = (X_1/m)/(X_2/n)$ where $X_1 = \chi^2(m)$ and $X_2 = \chi^2(n)$. Consequently, $1/W = (X_2/n)/(X_1/m)$, which is $F(n,m)$.

4. Suppose we want $P\{W \le c\} = .05$. Equivalently, $P\{1/W \ge 1/c\} = .05$, hence $P\{1/W \le 1/c\} = .95$. By Problem 3, $1/W$ is $F(n,m)$, so $1/c$ can be found from the $F$ table, and we can then compute $c$. The analysis is similar for .1, .025 and .01.
5. If $N$ is normal (0,1), then $T(n) = N/(\sqrt{\chi^2(n)/n})$. Thus $T^2(n) = N^2/(\chi^2(n)/n)$. But $N^2$ is $\chi^2(1)$, and the result follows.
6. If $Y = 2X$ then $f_Y(y) = f_X(x)|dx/dy| = (1/2)e^{-x} = (1/2)e^{-y/2}, y \ge 0$, the chi-square density with two degrees of freedom. If $X_1$ and $X_2$ are independent exponential random variables, then

$$\frac{X_1}{X_2} = \frac{(2X_1)/2}{(2X_2)/2} = \frac{\chi^2(2)/2}{\chi^2(2)/2} = F(2,2).$$

## Lecture 6

1. Apply the formula for the joint density of $Y_j$ and $Y_k$ with $j = 1, k = 3, n = 3, F(x) = x, f(x) = 1, 0 < x < 1$. The result is $f_{Y_1 Y_3}(x,y) = 6(y-x), 0 < x < y < 1$. Now let $Z = Y_3 - Y_1, W = Y_3$. The Jacobian of the transformation has absolute value 1, so $f_{ZW}(z,w) = f_{Y_1 Y_3}(y_1, y_3) = 6(y_3 - y_1) = 6z, 0 < z < w < 1$. Thus

$$f_Z(z) = \int_{w=z}^{1} 6z\, dw = 6z(1-z), \quad 0 < z < 1.$$

2. The probability that more than one random variable falls in $[x, x+dx]$ need not be negligible. For example, there can be a positive probability that two observations coincide with $x$.
3. The density of $Y_k$ is

$$f_{Y_k}(x0 = \frac{n!}{(k-1)!(n-k)!} x^{k-1}(1-x)^{n-k}, \quad 0 < x < 1$$

which is beta with $\alpha = k$ and $\beta = n-k+1$. (Note that $\Gamma(k) = (k-1)!, \Gamma(n-k+1) = (n-k)!, \Gamma(k+n-k+1) = \Gamma(n+1) = n!$.)

4. We have $Y_k > p$ if and only if *at most* $k-1$ observations are in $[0,p]$. But the probability that a particular observation lies in $[0,p]$ is $p/1 = p$. Thus we have $n$ Bernoulli trials with probability of success $p$ on a given trial. Explicitly,

$$P\{Y_k > p\} = \sum_{i=0}^{k-1} \binom{n}{i} p^i (1-p)^{n-i}.$$

## Lecture 7

1. Let $W_n = (S_n - E(S_n))/n$; then $E(W_n) = 0$ for all $n$, and

$$\operatorname{Var} W_n = \frac{\operatorname{Var} S_n}{n^2} = \frac{1}{n^2} \sum_{i=1}^{n} \sigma_i^2 \le \frac{nM}{n^2} = \frac{M}{n} \to 0.$$

It follows that $W_n \xrightarrow{P} 0$.

2. All $X_i$ and $X$ have the same distribution ($p(1) = p(0) = 1/2$), so $X_n \xrightarrow{d} 0$. But if $0 < \epsilon < 1$ then $P\{|X_n - X| \geq \epsilon\} = P\{X_n \neq X\}$, which is 0 for $n$ odd and 1 for $n$ even. Therefore $P\{|X_n - X| \geq \epsilon\}$ oscillates and has no limit as $n \to \infty$.
3. By the weak law of large numbers, $\overline{X}_n$ converges in probability to $\mu$, hence converges in distribution to $\mu$. Thus we can take $X$ to have a distribution function $F$ that is degenerate at $\mu$, in other words,

$$F(x) = \begin{cases} 0, & x < \mu \\ 1, & x \geq \mu. \end{cases}$$

4. Let $F_n$ be the distribution function of $X_n$. For all $x$, $F_n(x) = 0$ for sufficiently large $n$. Since the identically zero function cannot be a distribution function, there is no limiting distribution.

## Lecture 8

1. Note that $M_{X_n} = 1/(1-\beta t)^n$ where $1/(1-\beta t)$ is the moment-generating function of an exponential random variable (which has mean $\beta$). By the weak law of large numbers, $X_n/n \xrightarrow{P} \beta$, hence $X_n/n \xrightarrow{d} \beta$.
2. $\chi^2(n) = \sum_{i=1}^n X_i^2$, where the $X_i$ are iid, each normal (0,1). Thus the central limit theorem applies.
3. We have $n$ Bernoulli trials, with probability of success $p = \int_a^b f(x)\,dx$ on a given trial. Thus $Y_n$ is binomial $(n, p)$. If $n$ and $p$ satisfy the sufficient condition given in the text, the normal approximation with $E(Y_n) = np$ and $\operatorname{Var} Y_n = np(1-p)$ should work well in practice.
4. We have $E(X_i) = 0$ and

$$\operatorname{Var} X_i = E(X_i^2) = \int_{-1/2}^{1/2} x^2\,dx = 2\int_0^{1/2} x^2\,dx = 1/12.$$

By the central limit theorem, $Y_n$ is approximately normal with $E(Y_n) = 0$ and $\operatorname{Var} Y_n = n/12$.

5. Let $W_n = n(1 - F(Y_n))$. Then

$$P\{W_n \geq w\} = P\{F(Y_n) \leq 1 - (w/n)\} = P\{\max F(X_i) \leq 1 - (w/n)\}$$

hence

$$P\{W_n \geq w\} = \left(1 - \frac{w}{n}\right)^n, \quad 0 \leq w \leq n,$$

which approaches $e^{-w}$ as $n \to \infty$. Therefore the limiting distribution of $W_n$ is exponential.

## Lecture 9

1. (a) We have

$$f_\theta(x_1,\dots,x_n)=\theta^{x_1+\cdots+x_n}\frac{e^{-n\theta}}{x_1!\cdots x_n!}.$$

With $x=x_1+\cdots+x_n$, take logarithms and differentiate to get

$$\frac{\partial}{\partial\theta}(x\ln\theta-n\theta)=\frac{x}{\theta}-n=0,\quad \hat{\theta}=\overline{X}.$$

(b) $f_\theta(x_1,\dots,x_n)=\theta^n(x_1\cdots x_n)^{\theta-1},\theta>0$, and

$$\frac{\partial}{\partial\theta}(n\ln\theta+(\theta-1)\sum_{i=1}^n\ln x_i)=\frac{n}{\theta}+\sum_{i=1}^n\ln x_i=0,\quad \hat{\theta}=-\frac{n}{\sum_{i=1}^n}\ln x_i.$$

Note that $0<x_i<1$, so $\ln x_i<0$ for all $i$ and $\hat{\theta}>0$.

(c) $f_\theta(x_1,\dots,x_n)=(1/\theta^n)\exp[-(\sum_{i=1}^n x_i)/\theta]$. With $x=\sum_{i=1}^n x_i$ we have

$$\frac{\partial}{\partial\theta}(-n\ln\theta-\frac{x}{\theta})=-\frac{n}{\theta}+\frac{x}{\theta^2}=0,\quad \hat{\theta}=\overline{X}$$

(d) $f_\theta(x_1,\dots,x_n)=(1/2)^n\exp[-\sum_{i=1}^n|x_i-\theta|]$. We must minimize $\sum_{i=1}^n|x_i-\theta|$, and we must be careful when differentiating because of the absolute values. If the order statistics of the $x_i$ are $y_i, i=1,\dots,n$, and $y_k<\theta<y_{k+1}$, then the sum to be minimized is

$$(\theta-y_1)+\cdots+(\theta-y_k)+(y_{k+1}-\theta)+\cdots+(y_n-\theta).$$

The derivative of the sum is the number of $y_i$'s less than $\theta$ minus the number of $y_i$'s greater than $\theta$. Thus as $\theta$ increases, $\sum_{i=1}^n|x_i-\theta|$ decreases until the number of $y_i$'s less than $\theta$ equals the number of $y_i$'s greater than $\theta$. We conclude that $\hat{\theta}$ is the *median* of the $X_i$.

(e) $f_\theta(x_1,\dots,x_n)=\exp[-\sum_{i=1}^n x_i]e^{n\theta}$ if all $x_i\ge\theta$, and 0 elsewhere. Thus

$$f_\theta(x_1,\dots,x_n)=\exp[-\sum_{i=1}^n x_i]e^{n\theta}I[\theta\le\min(x_1,\dots,x_n)].$$

The indicator $I$ prevents us from differentiating blindly. As $\theta$ increases, so does $e^{n\theta}$, but if $\theta>\min_i x_i$, the indicator drops to 0. Thus $\hat{\theta}=\min(X_1,\dots,X_n)$.

2. $f_\theta(x_1,\dots,x_n)=1$ if $\theta-(1/2)\le x_i\le\theta+(1/2)$ for all $i$, and 0 elsewhere. If $Y_1,\dots,Y_n$ are the order statistics of the $X_i$, then $f_\theta(x_1,\dots,x_n)=I[y_n-(1/2)\le\theta\le y_1+(1/2)]$, where $y_1=\min x_i$ and $y_n=\max x_i$. Thus any function $h(X_1,\dots,X_n)$ such that

$$Y_n-\frac{1}{2}\le h(X_1,\dots,X_n)\le Y_1+\frac{1}{2}$$

for all $X_1, \dots, X_n$) is an MLE of $\theta$. Some solutions are $h = Y_1 + (1/2)$, $h = Y_n - (1/2)$, $h = (Y_1 + Y_n)/2$, $h = (2Y_1 + 4Y_n - 1)/6$ and $h = (4Y_1 + 2Y_n + 1)/6$. In all cases, the inequalities reduce to $Y_n - Y_1 \leq 1$, which is true.

3. (a) $X_i$ is Poisson $(\theta)$ so $E(X_i) = \theta$. The method of moments sets $\overline{X} = \theta$, so the estimate of $\theta$ is $\theta^* = \overline{X}$, which is consistent by the weak law of large numbers.

   (b) $E(X_i) = \int_0^1 \theta x^\theta \, d\theta = \theta/(\theta+1) = \overline{X}, \theta = \theta\overline{X} + \overline{X}$, so

$$\theta^* = \frac{\overline{X}}{1-\overline{X}} \xrightarrow{P} \frac{\theta/(\theta+1)}{1-[\theta/(\theta+1)]} = \theta$$

   hence $\theta^*$ is consistent.

   (c) $E(X_i) = \theta = \overline{X}$, so $\theta^* = \overline{X}$, consistent by the weak law of large numbers.

   (d) By symmetry, $E(X_i) = \theta$ so $\theta^* = \overline{X}$ as in (a) and (c).

   (e) $E(X_i) = \int_\theta^\infty x e^{-(x-\theta)} \, dx =$ (with $y = x - \theta$) $\int_0^\infty (y+\theta)e^{-y} \, dy = 1 + \theta = \overline{X}$. Thus $\theta^* = \overline{X} - 1$ which converges in probability to $(1+\theta) - 1 = \theta$, proving consistency.

4. $P\{X \leq r\} = \int_0^r (1/\theta)e^{-x/\theta} \, dx = \left[-e^{-x/\theta}\right]_0^r = 1 - e^{-r/\theta}$. The MLE of $\theta$ is $\hat{\theta} = \overline{X}$ [see Problem 1(c)], so the MLE of $1 - e^{-r/\theta}$ is $1 - e^{-r/\overline{X}}$.

5. The MLE of $\theta$ is $X/n$, the relative frequency of success. Since

$$P\{a \leq X \leq b\} = \sum_{k=a}^{b} \binom{n}{k} \theta^k (1-\theta)^{n-k},$$

   the MLE of $P\{a \leq X \leq b\}$ is found by replacing $\theta$ by $X/n$ in the above summation.

## Lecture 10

1. Set $2\Phi(b) - 1$ equal to the desired confidence level. This, along with the table of the normal (0,1) distribution function, determines $b$. The length of the confidence interval is $2b\sigma/\sqrt{n}$.

2. Set $2F_T(b) - 1$ equal to the desired confidence level. This, along with the table of the $T(n-1)$ distribution function, determines $b$. The length of the confidence interval is $2bS/\sqrt{n-1}$.

3. In order to compute the expected length of the confidence interval, we must compute $E(S)$, and the key observation is

$$S = \frac{\sigma}{\sqrt{n}} \sqrt{\frac{nS^2}{\sigma^2}} = \frac{\sigma}{\sqrt{n}} \sqrt{\chi^2(n-1)}.$$

   If $f(x)$ is the chi-square density with $r = n - 1$ degrees of freedom [see (3.8)], then the expected length is

$$\frac{2b}{\sqrt{n-1}} \frac{\sigma}{\sqrt{n}} \int_0^\infty x^{1/2} f(x) \, dx$$

   and an appropriate change of variable reduces the integral to a gamma function which can be evaluated explicitly.

4. We have $E(X_i) = \alpha\beta$ and $\text{Var}(X_i) = \alpha\beta^2$. For large $n$,

$$\frac{\overline{X} - \alpha\beta}{\sqrt{\alpha}\beta/\sqrt{n}} = \frac{\overline{X} - \mu}{\mu/\sqrt{\alpha n}}$$

is approximately normal (0,1) by the central limit theorem. With $c = 1/\sqrt{\alpha n}$ we have

$$P\{-b < \frac{\overline{X} - \mu}{c\mu} < b\} = \Phi(b) - \Phi(-b) = 2\Phi(b) - 1$$

and if we set this equal to the desired level of confidence, then $b$ is determined. The confidence interval is given by $(1 - bc)\mu < \overline{X} < (1 + bc)\mu$, or

$$\frac{\overline{X}}{1 + bc} < \mu < \frac{\overline{X}}{1 - bc}$$

where $c \to 0$ as $n \to \infty$.

5. A confidence interval of length $L$ corresponds to $|(Y_n/n) - p| < L/2$, an event with probability

$$2\Phi\left(\frac{L\sqrt{n}/2}{\sqrt{p(1-p)}}\right) - 1.$$

Setting this probability equal to the desired confidence level gives an inequality of the form

$$\frac{L\sqrt{n}/2}{\sqrt{p(1-p)}} > c.$$

As in the text, we can replace $p(1-p)$ by its maximum value $1/4$. We find the minimum value of $n$ by squaring both sides.

In the first example in (10.1), we have $L = .02, L/2 = .01$ and $c = 1.96$. This problem essentially reproduces the analysis in the text in a more abstract form. Specifying how close to $p$ we want our estimate to be (at the desired level of confidence) is equivalent to specifying the length of the confidence interval.

## Lecture 11

1. Proceed as in (11.1):

$$Z = \overline{X} - \overline{Y} - (\mu_1 - \mu_2) \quad \text{divided by} \quad \sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}}$$

is normal (0,1), and $W = (nS_1^2/\sigma_1^2) + (mS_2^2/\sigma_2^2)$ is $\chi^2(n+m-2)$. Thus $\sqrt{n+m-2}Z/\sqrt{W}$ is $T(n+m-2)$, but the unknown variances cannot be eliminated.

2. If $\sigma_1^2 = c\sigma_2^2$, then

$$\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m} = c\sigma_2^2(\frac{1}{n} + \frac{1}{cm})$$

and

$$\frac{nS_1^2}{\sigma_1^2} + \frac{mS_2^2}{\sigma_2^2} = \frac{nS_1^2 + cmS_2^2}{c\sigma_2^2}.$$

Thus $\sigma_2^2$ can again be eliminated, and confidence intervals can be constructed, assuming $c$ known.

## Lecture 12

1. The given test is an LRT and is completely determined by $c$, independent of $\theta > \theta_0$.
2. The likelihood ratio is $L(x) = f_1(x)/f_0(x) = (1/4)/(1/6) = 3/2$ for $x = 1, 2$, and $L(x) = (1/8)/(1/6) = 3/4$ for $x = 3, 4, 5, 6$. If $0 \le \lambda < 3/4$, we reject for all $x$, and $\alpha = 1, \beta = 0$. If $3/4 < \lambda < 3/2$, we reject for $x = 1, 2$ and accept for $x = 3, 4, 5, 6$, with $\alpha = 1/3$ and $\beta = 1/2$. If $3/2 < \lambda \le \infty$, we accept for all $x$, with $\alpha = 0, \beta = 1$.

   For $\alpha = .1$, set $\lambda = 3/2$, accept when $x = 3, 4, 5, 6$, reject with probability $a$ when $x = 1, 2$. Then $\alpha = (1/3)a = .1, a = .3$ and $\beta = (1/2) + (1/2)(1 - a) = .85$.
3. Since (220-200)/10=2, it follows that when $c$ reaches 2, the null hypothesis is accepted. The associated type 1 error probability is $\alpha = 1 - \Phi(2) = 1 - .977 = .023$. Thus the given result is significant even at the significance level .023. If we were to take additional observations, enough to drive the probability of a type 1 error down to .023, we would still reject $H_0$. Thus the $p$-value is a concise way of conveying a lot of information about the test.

## Lecture 13

1. We sum $(X_i - np_i)^2/np_i, i = 1, 2, 3$, where the $X_i$ are the observed frequencies and the $np_i = 50, 30, 20$ are the expected frequencies. The chi-square statistic is

$$\frac{(40-50)^2}{50} + \frac{(33-30)^2}{30} + \frac{(27-20)^2}{20} = 2 + .3 + 2.45 = 4.75$$

   Since $P\{\chi^2(2) > 5.99\} = .05$ and $4.75 < 5.99$, we accept $H_0$.
2. The expected frequencies are given by

| | $A$ | $B$ | $C$ |
|---|---|---|---|
| 1 | 49 | 147 | 98 |
| 2 | 51 | 153 | 102 |

   For example, to find the entry in the 2C position, we can multiply the row 2 sum by the column 3 sum and divide by the total number of observations (namely 600) to get

(306)(200)/600=102. Alternatively, we can compute $P(C) = (98 + 102)/600 = 1/3$. We multiply this by the row 2 sum 306 to get 306/3=102. The chi-square statistic is

$$\frac{(33-49)^2}{49} + \frac{(147-147)^2}{147} + \frac{(114-98)^2}{98} + \frac{(67-51)^2}{51} + \frac{(153-153)^2}{153} + \frac{(86-102)^2}{102}$$

which is 5.224+0+2.612+5.02+0+2.510 = 15.366. There are $(h-1)(k-1) = 1\times 2 = 2$ degrees of freedom, and $P\{\chi^2(2) > 5.99\} = .05$. Since $15.366 > 5.94$, we reject $H_0$.

3. The observed frequencies minus the expected frequencies are

$$a - \frac{(a+b)(a+c)}{a+b+c+d} = \frac{ad-bc}{a+b+c+d}, \quad b - \frac{(a+b)(b+d)}{a+b+c+d} = \frac{bc-ad}{a+b+c+d},$$

$$c - \frac{(a+c)(c+d)}{a+b+c+d} = \frac{bc-ad}{a+b+c+d}, \quad d - \frac{(c+d)(b+d)}{a+b+c+d} = \frac{ad-bc}{a+b+c+d}.$$

The chi-square statistic is

$$\frac{(ad-bc)^2}{a+b+c+d}\left[\frac{1}{(a+b)(c+d)(a+c)(b+d)}\right]\times$$

$$[(c+d)(b+d) + (a+c)(c+d) + (a+b)(b+d) + (a+b)(a+c)]$$

and the expression in small brackets simplifies to $(a+b+c+d)^2$, and the result follows.

## Lecture 14

1. The joint probability function is

$$f_\theta(x_1,\dots,x_n) = \prod_{i=1}^n \frac{e^{-\theta\theta^{x_i}}}{x_i!} = \frac{e^{-n\theta}\theta^{u(x)}}{x_1!\cdots x_n!}.$$

Take $g(\theta, u(x)) = e^{-n\theta}\theta^{u(x)}$ and $h(x) = 1/(x_1!\cdots x_n!)$.

2. $f_\theta(x_1,\dots,x_n) = [A(\theta)]^n B(x_1)\cdots B(x_n)$ if $0 < x_i < \theta$ for all $i$, and 0 elsewhere. This can be written as

$$[A(\theta)]^n \prod_{i=1}^n B(x_i) I\big[\max_{1\le i\le n} x_i < \theta\big]$$

where $I$ is an indicator. We take $g(\theta, u(x)) = A^n(\theta) I[\max x_i < \theta]$ and $h(x) = \prod_{i=1}^n B(x_i)$.

3. $f_\theta(x_1,\dots,x_n) = \theta^n(1-\theta)^{u(x)}$, and the factorization theorem applies with $h(x) = 1$.

4. $f_\theta(x_1,\dots,x_n) = \theta^{-n}\exp[-(\sum_{i=1}^n x_i)/\theta]$, and the factorization theorem applies with $h(x) = 1$.

5. $f_\theta(x) = \Gamma(a+b)/[\Gamma(a)\Gamma(b)]x^{a-1}(1-x)^{b-1}$ on (0,1). In this case, $a = \theta$ and $b = 2$. Thus $f_\theta(x) = (\theta+1)\theta x^{\theta-1}(1-x)$, so

$$f_\theta(x_1, \dots, x_n) = (\theta+1)^n \theta^n \big(\prod_{i=1}^n x_i\big)^{\theta-1} \prod_{i=1}^n (1-x_i)$$

and the factorization theorem applies with

$$g(\theta, u(x)) = (\theta+1)^n \theta^n u(x)^{\theta-1}$$

and $h(x) = \prod_{i=1}^n (1-x_i)$.

6. $f_\theta(x) = (1/[\Gamma(\alpha)\beta^\alpha])x^{\alpha-1}e^{-x/\beta}, x > 0$, with $\alpha = \theta$ and $\beta$ arbitrary. The joint density is

$$f_\theta(x_1, \dots, x_n) = \frac{1}{[\Gamma(\alpha)]^n \beta^{n\alpha}} u(x)^{\alpha-1} \exp[-\sum_{i=1}^n x_i/\beta]$$

and the factorization theorem applies with $h(x) = \exp[-\sum x_i/\beta]$ and $g(\theta, u(x))$ equal to the remaining factors.

7. We have

$$P_\theta\{X'_1 = x_1, \dots, X'_n = x_n\} = P_\theta\{Y = y\} P\{X_1 = x_1, \dots, X_n = x_n | Y = y\}$$

We can drop the subscript $\theta$ since $Y$ is sufficient, and we can replace $X'_i$ by $X_i$ by definition of B's experiment. The result is

$$P_\theta\{X'_1 = x_1, \dots, X'_n = x_n\} = P_\theta\{X_1 = x_1, \dots, X_n = x_n\}$$

as desired.

## Lecture 17

1. Take $u(X) = X$.
2. The joint density is

$$f_\theta(x_1, \dots, x_n) = \exp\big[-\sum_{i=1}^n (x_i - \theta)\big] I[\min x_i > \theta]$$

so $Y_1$ is sufficient. Now if $y > \theta$, then

$$P\{Y_1 > y\} = (P\{X_1 > y\})^n = \Big(\int_y^\infty \exp[-(x-\theta)]\, dx\Big)^n = \exp[-n(y-\theta)],$$

so

$$F_{Y_1}(y) = 1 - e^{-n(y-\theta)}, \quad f_{Y_1}(y) = ne^{-n(y-\theta)}, \quad y > \theta.$$

The expectation of $g(Y_1)$ under $\theta$ is

$$E_\theta[g(Y_1)] = \int_\theta^\infty g(y) n \exp[-n(y-\theta)]\, dy.$$

If this is 0 for all $\theta$, divide by $e^{n\theta}$ to get

$$\int_\theta^\infty g(y) n \exp(-ny)\, dy = 0.$$

Differentiating with respect to $\theta$, we have $-g(\theta) n \exp(-n\theta) = 0$, so $g(\theta) = 0$ for all $\theta$, proving completeness. The expectation of $Y_1$ under $\theta$ is

$$\int_\theta^\infty yn \exp[-n(y-\theta)]\, dy = \int_\theta^\infty (y-\theta) n \exp[-n(y-\theta)]\, dy + \theta \int_\theta^\infty n \exp[-n(y-\theta)]\, dy$$

$$= \int_0^\infty zn \exp(-nz)\, dz + \theta = \frac{1}{n} + \theta.$$

Thus $E_\theta[Y_1 - (1/n)] = \theta$, so $Y_1 - (1/n)$ is a UMVUE of $\theta$.

3. Since $f_\theta(x) = \theta \exp[(\theta - 1)\ln x]$, the density belongs to the exponential class. Thus $\sum_{i=1}^n \ln X_i$ is a complete sufficient statistic, hence so is $\exp\left[(1/n)\sum_{i=1}^n \ln X_i\right] = u(X_1, \dots, X_n)$. The key observation is that if $Y$ is sufficient and $g$ is one-to-one, then $g(Y)a$ is also sufficient, since $g(Y)$ conveys exactly the same information as $Y$ does; similarly for completeness.

   To compute the maximum likelihood estimate, note that the joint density is $f_\theta(x_1, \dots, x_n) = \theta^n \exp[(\theta-1)\sum_{i=1}^n \ln x_i]$. Take logarithms, differentiate with respect to $\theta$, and set the result equal to 0. We get $\hat\theta = -n/\sum_{i=1}^n \ln X_i$, which is a function of $u(X_1, \dots, X_n)$.

4. Each $X_i$ is gamma with $\alpha = 2, \beta = 1/\theta$, so (see Lecture 3) $Y$ is gamma $(2n, 1/\theta)$. Thus

$$E_\theta(1/Y) = \int_0^\infty (1/y) \frac{1}{\Gamma(2n)(1/\theta)^{2n}} y^{2n-1} e^{-\theta y}\, dy$$

   which becomes, under the change of variable $z = \theta y$,

$$\frac{\theta^{2n}}{\Gamma(2n)} \int_0^\infty \frac{z^{2n-2}}{\theta^{2n-2}} e^{-z} \frac{dz}{\theta} = \frac{\theta^{2n}}{\theta^{2n-1}} \frac{\Gamma(2n-1)}{\Gamma(2n)} = \frac{\theta}{2n-1}.$$

   Therefore $E_\theta[(2n-1)/Y] = \theta$, and $(2n-1)/Y$ is the UMVUE of $\theta$.

5. We have $E(Y_2) = [E(X_1) + E(X_2)]/2 = \theta$, hence $E[E(Y_2|Y_1)] = E(Y_2) = \theta$. By completeness, $E(Y_2|Y_1)$ must be $Y_1/n$.

6. Since $X_i/\sqrt{\theta}$ is normal (0,1), $Y/\theta$ is $\chi^2(n)$, which has mean $n$ and variance $2n$. Thus $E[(Y/\theta)^2] = n^2 + 2n$, so $E(Y^2) = \theta^2(n^2+2n)$.Therefore the UMVUE of $\theta^2$) is $Y^2/(n^2 + 2n)$.

7. (a) $E[E(I|Y)] = E(I) = P\{X_1 \le 1\}$, and the result follows by completeness.
(b) We compute

$$P\{X_1 = r|X_1 + \cdots + X_n = s\} = \frac{P\{X_1 = r, X_2 + \cdots + X_n = s - r\}}{P\{X_1 + \cdots + X_n\} = s}.$$

The numerator is

$$\frac{e^{-\theta}\theta^r}{r!} e^{-(n-1)\theta} \frac{[(n-1)\theta]^{s-r}}{(s-r)!}$$

and the denominator is

$$\frac{e^{-n\theta}(n\theta)^s}{s!}$$

so the conditional probability is

$$\binom{s}{r} \frac{(n-1)^{s-r}}{n^s} = \binom{s}{r} \left(\frac{n-1}{n}\right)^{s-r} \left(\frac{1}{n}\right)^r$$

which is the probability of $r$ successes in $s$ Bernoulli trials, with probability of success $1/n$ on a given trial. Intuitively, if the sum is $s$, then each contribution to the sum is equally likely to come from $X_1, \ldots, X_n$.
(c) By (b), $P\{X_1 = 0|Y\} + P\{X_1 = 1|Y\}$ is given by

$$\left(1 - \frac{1}{n}\right)^Y + Y\left(\frac{1}{n}\right)\left(1 - \frac{1}{n}\right)^{Y-1} = \left(\frac{n-1}{n}\right)^Y \left[1 + \frac{Y/n}{(n-1)/n}\right]$$

$$= \left(\frac{n-1}{n}\right)^Y \left[1 + \frac{Y}{n-1}\right].$$

This formula also works for $Y = 0$ because it evaluates to 1.

8. The joint density is

$$f_\theta(x_1, \ldots, x_n) = \frac{1}{\theta_2^n} \exp\left[-\sum_{i=1}^n \frac{(x_i - \theta_1)}{\theta_2}\right] I\big[\min_i X_i > \theta_1\big].$$

Since

$$\sum_{i=1}^n \frac{(x_i - \theta_1)}{\theta_2} = \frac{1}{\theta_2} \sum_{i=1}^n x_i - n\theta_1,$$

the result follows from the factorization theorem.

## Lecture 18

1. By (18.4), the numerator of $\delta(x)$ is

$$\int_0^1 \theta\theta^{r-1}(1-\theta)^{s-1}\binom{n}{x}\theta^x(1-\theta)^{n-x}\,d\theta$$

and the denominator is

$$\int_0^1 \theta^{r-1}(1-\theta)^{s-1}\binom{n}{x}\theta^x(1-\theta)^{n-x}\,d\theta.$$

Thus $\delta(x)$ is

$$\frac{\beta(r+x+1,n-x+s)}{\beta(r+x,n-x+s)}=\frac{\Gamma(r+x+1)}{\Gamma(r+x)}\frac{\Gamma(r+s+n)}{\Gamma(r+s+n+1)}=\frac{r+x}{r+s+n}.$$

2. The risk function is

$$E_\theta\left[\left(\frac{r+X}{r+s+n}-\theta\right)^2\right]=\frac{1}{(r+s+n)^2}E_\theta[(X-n\theta+r-r\theta-s\theta)^2]$$

with $E_\theta(X-n\theta)=0, E_\theta[(X-n\theta)^2=\operatorname{Var}X=n\theta(1-\theta)$. Thus

$$R_\delta(\theta)=\frac{1}{(r+s+n)^2}[n\theta(1-\theta)+(r-r\theta-s\theta)^2].$$

The quantity in brackets is

$$n\theta-n\theta^2+r^2+r^2\theta^2+s^2\theta^2-2r^2\theta-2rs\theta+2rs\theta^2$$

which simplifies to

$$((r+s)^2-n)\theta^2+(n-2r(r+s))\theta+r^2$$

and the result follows.

3. If $r=s=\sqrt{n}/2$, then $(r+s)^2-n=0$ and $n-2r(r+s)=0$, so

$$R_\delta(\theta)=\frac{r^2}{(r+s+n)^2}.$$

4. The average loss using $\delta$ is $B(\delta)=\int_{-\infty}^{\infty}h(\theta)R_\delta(\theta)\,d\theta$. If $\psi(x)$ has a smaller maximum risk than $\delta(x)$, then since $R_\delta$ is constant, we have $R_\psi(\theta)<R_\delta(\theta)$ for all $\theta$. Therefore $B(\psi)<B(\delta)$, contradicting the fact that $\theta$ is a Bayes estimate.

## Lecture 20

1.

$$\operatorname{Var}(XY)=E[(XY)^2]-(EXEY)^2=E(X^2)E(Y^2)-(EX)^2(EY)^2$$

$$(\sigma_X^2+\mu_X^2)(\sigma_Y^2+\mu_Y^2)-\mu_X^2\mu_Y^2=\sigma_X^2\sigma_Y^2+\mu_X^2\sigma_Y^2+\mu_Y^2\sigma_X^2.$$

2.

$$\operatorname{Var}(aX+bY) = \operatorname{Var}(aX) + \operatorname{Var}(bY) + 2ab\operatorname{Cov}(X,Y)$$

$$= a^2\sigma_X^2 + b^2\sigma_Y^2 + 2ab\rho\sigma_X\sigma_Y.$$

3.

$$\operatorname{Cov}(X, X+Y) = \operatorname{Cov}(X,X) + \operatorname{Cov}(X,Y) = \operatorname{Var} X + 0 = \sigma_X^2.$$

4. By Problem 3,

$$\rho_{X,X+Y} = \frac{\sigma_X^2}{\sigma_X\sigma_{X+Y}} = \frac{\sigma_X}{\sqrt{\sigma_X^2+\sigma_Y^2}}.$$

5.

$$\operatorname{Cov}(XY, X) = E(X^2)E(Y) - E(X)^2E(Y)$$

$$= (\sigma_X^2 + \mu_X^2)\mu_Y - \mu_X^2\mu_Y = \sigma_X^2\mu_Y.$$

6. We can assume without loss of generality that $E(X^2) > 0$ and $E(Y^2) > 0$. We will have equality iff the discriminant $b^2 - 4ac = 0$, which holds iff $h(\lambda) = 0$ for some $\lambda$. Equivalently, $\lambda X + Y = 0$ for some $\lambda$. We conclude that equality holds if and only if $X$ and $Y$ are *linearly* dependent.

## Lecture 21

1. Let $Y_i = X_i - E(X_i)$; then $E[\big(\sum_{i=1}^n t_iY_i\big)^2] \geq 0$ for all $\underline{t}$. But this expectation is

$$E[\sum_i t_iY_i \sum_j t_jY_j] = \sum_{i,j} t_i\sigma_{ij}t_j = \underline{t}'K\underline{t}$$

where $\sigma_{ij} = \operatorname{Cov}(X_i, X_j)$. By definition of covariance, $K$ is symmetric, and $K$ is always *nonnegative definite* because $\underline{t}'K\underline{t} \geq 0$ for all $\underline{t}$. Thus all eigenvalues $\lambda_i$ of $K$ are nonnegative. But $K = LDL'$, so $\det K = \det D = \lambda_1 \cdots \lambda_n$. If $K$ is nonsingular then all $\lambda_i > 0$ and $K$ is positive definite.

2. We have $\underline{X} = C\underline{Z} + \underline{\mu}$ where $C$ is nonsingular and the $Z_i$ are independent normal random variables with zero mean. Then $\underline{Y} = A\underline{X} = AC\underline{Z} + A\underline{\mu}$, which is Gaussian.

3. The moment-generating function of $(X_1, \dots, X_m)$ is the moment-generating function of $(X_1, \dots, X_n)$ with $t_{m+1} = \cdots = t_n = 0$. We recognize the latter moment-generating function as Gaussian; see (21.1).

4. Let $Y = \sum_{i=1}^n c_iX_i$; then

$$E(e^{tY}) = E\big[\exp\big(\sum_{i=1}^n c_itX_i\big)\big] = M_{\underline{X}}(c_1t, \dots, c_nt)$$

$$= \exp\big(t\sum_{i=1}^{n} c_i\mu_i\big)\exp\big(\frac{1}{2}t^2\sum_{i,j=1}^{n} c_i a_{ij} c_j\big)$$

which is the moment-generating function of a normally distributed random variable. Another method: Let $W = c_1X_1 + \cdots + c_nX_n = \underline{c}'\underline{X} = \underline{c}'(A\underline{Y} + \underline{\mu})$, where the $Y_i$ are independent normal random variables with zero mean. Thus $W = \underline{b}'\underline{Y} + \underline{c}'\underline{\mu}$ where $\underline{b}' = \underline{c}'A$. But $\underline{b}'\underline{Y}$ is a linear combination of independent normal random variables, hence is normal.

## Lecture 22

1. If $y$ is the best estimate of $Y$ given $X = x$, then

$$y - \mu_Y = \frac{\rho\sigma_Y}{\sigma_X}(x - \mu_X)$$

and [see (20.1)] the minimum mean square error is $\sigma_Y^2(1-\rho^2)$, which in this case is 28. We are given that $\rho\sigma_Y/\sigma_X = 3$, so $\rho\sigma_Y = 3 \times 2 = 6$ and $\rho^2 = 36/\sigma_Y^2$. Therefore

$$\sigma_Y^2(1 - \frac{36}{\sigma_Y^2}) = \sigma_Y^2 - 36 = 28, \quad \sigma_Y = 8, \quad \rho^2 = \frac{36}{64}, \quad \rho = .75.$$

Finally, $y = \mu_Y + 3x - 3\mu_X = \mu_Y + 3x + 3 = 3x + 7$, so $\mu_Y = 4$.

2. The bivariate normal density is of the form

$$f_\theta(x,y) = a(\theta)b(x,y)\exp[p_1(\theta)x^2 + p_2(\theta)y^2 + p_3(\theta)xy + p_4(\theta)x + p_5(\theta)y]$$

so we are in the exponential class. Thus

$$\big(\sum X_i^2, \quad \sum Y_i^2, \quad \sum X_iY_i, \quad \sum X_i, \quad \sum Y_i\big)$$

is a complete sufficient statistic for $\theta = (\sigma_X^2, \sigma_Y^2, \rho, \mu_X, \mu_Y)$. Note also that any statistic in one-to-one correspondence with this one is also complete and sufficient.

## Lecture 23

1. The probability of any event is found by integrating the density on the set defined by the event. Thus

$$P\{a \le f(X) \le b\} = \int_A f(x)\,dx, \quad A = \{x : a \le f(x) \le b\}.$$

2. Bernoulli: $f_\theta(x) = \theta^x(1-\theta)^{1-x}, x = 0, 1$

$$\frac{\partial}{\partial\theta}\ln f_\theta(x) = \frac{\partial}{\partial\theta}[x\ln\theta + (1-x)\ln(1-\theta)] = \frac{x}{\theta} - \frac{1-x}{1-\theta}$$

$$\frac{\partial^2}{\partial\theta^2}\ln f_\theta(x) = -\frac{x}{\theta^2} - \frac{1-x}{(1-\theta)^2}$$

$$I(\theta) = E_\theta\big[\frac{X}{\theta^2} + \frac{1-X}{(1-\theta)^2}\big] = \frac{1}{\theta} + \frac{1}{1-\theta} = \frac{1}{\theta(1-\theta)}$$

since $E_\theta(X) = \theta$. Now

$$\operatorname{Var}_\theta Y \geq \frac{1}{nI(\theta)} = \frac{\theta(1-\theta)}{n}.$$

But

$$\operatorname{Var}_\theta \overline{X} = \frac{1}{n^2}\operatorname{Var}[\text{binomial}(n,\theta)] = \frac{n\theta(1-\theta)}{n^2} = \frac{\theta(1-\theta)}{n}$$

so $\overline{X}$ is a UMVUE of $\theta$.

Normal:

$$f_\theta(x) = \frac{1}{\sqrt{2\pi}\sigma}\exp[-(x-\theta)^2/2\sigma^2]$$

$$\frac{\partial}{\partial\theta}\ln f_\theta(x) = \frac{\partial}{\partial\theta}\big[-\frac{(x-\theta)^2}{2\sigma^2}\big] = \frac{x-\theta}{\sigma^2}$$

$$\frac{\partial^2}{\partial\theta^2}\ln f_\theta(x) = -\frac{1}{\sigma^2}, \quad I(\theta) = \frac{1}{\sigma^2}, \quad \operatorname{Var}_\theta Y \geq \frac{\sigma^2}{n}$$

But $\operatorname{Var}_\theta \overline{X} = \sigma^2/n$, so $\overline{X}$ is a UMVUE of $\theta$.

Poisson: $f_\theta(x) = e^{-\theta}\theta^x/x!, x = 0, 1, 2\ldots$

$$\frac{\partial}{\partial\theta}\ln f_\theta(x) = \frac{\partial}{\partial\theta}(-\theta + x\ln\theta) = -1 + \frac{x}{\theta}$$

$$\frac{\partial^2}{\partial\theta^2}\ln f_\theta(x) = -\frac{x}{\theta^2}, \quad I(\theta) = E\big(\frac{X}{\theta^2}\big) = \frac{\theta}{\theta^2} = \frac{1}{\theta}$$

$$\operatorname{Var}_\theta Y \geq \frac{\theta}{n} = \operatorname{Var}_\theta \overline{X}$$

so $\overline{X}$ is a UMVUE of $\theta$.

## Lecture 25

1.

$$K(p) = \sum_{k=0}^{c} \binom{n}{k} p^k (1-p)^{n-k}$$

with $c = 2$ and $p = 1/2$ under $H_0$. Therefore

$$\alpha = \left[\binom{12}{0} + \binom{12}{1} + \binom{12}{2}\right](1/2)^n = \frac{79}{4096} = .019.$$

2. The deviations, with ranked absolute values in parentheses, are

   16.9(14), -1.7(5), -7.9(9), -1.2(4), 12.4(12), 9.8(10), -.3(2), 2.7(6), -3.4(7), 14.5(13), 24.4(16), 5.2(8), -12.2(11), 17.8(15), .1(1), .5(3)

   The Wilcoxon statistic is $W$=1-2+3-4-5+6-7+8-9+10-11+12+13+14+15+16=60

   Under $H_0$, $E(W) = 0$ and $\operatorname{Var} W = n(n+1)(2n+1)/6 = 1496$, $\quad \sigma_W = 38.678$

   Now $W/38.678$ is approximately normal (0,1) and $P\{W \geq c\} = P\{W/38.678 \geq c/38.678\} = .05$. From a normal table, $c/38.678 = 1.645$, $\quad c = 63.626$. Since $60 < 63.626$, we accept $H_0$.

3. The moment-generating function of $V_j$ is $M_{V_j}(t) = (1/2)(e^{-jt}+e^{jt})$ and the moment-generating function of $W$ is $M_W(t) = \prod_{j=1}^{n} M_{V_j}(t)$. When $n = 1$, $W = \pm 1$ with equal probability. When $n = 2$,

$$M_W(t) = \frac{1}{2}(e^{-t} + e^{t})\frac{1}{2}(e^{-2t} + e^{2t}) = \frac{1}{4}(e^{-3t} + e^{-t} + e^{t} + e^{3t})$$

   so $W$ takes on the values $-3, -1, 1, 3$ with equal probability. When $n = 3$,

$$M_W(t) = \frac{1}{4}(e^{-3t} + e^{-t} + e^{t} + e^{3t})\frac{1}{2}(e^{-3t} + e^{3t})$$

$$= \frac{1}{8}(e^{-6t} + e^{-4t} + e^{-2t} + 1 + 1 + e^{2t} + e^{4t} + e^{6t}).$$

   Therefoe $P\{W = k\} = 1/8$ for $k = -6, -4, -2, 2, 4, 6$, $P\{W = 0\} = 1/4$, and $P\{W = k\} = 0$ for other values of $k$