Statistics 582, Final Exam Solutions

Wellner; 3/12/2018

1. (30 points) **Define** *any three* of the following terms. In each case, provide an appropriate context for your definition.

   (a) A level α *permutation test* (in the setting of $m + n = N$ observations $X_1, \ldots, X_m, Y_1, \ldots, Y_n$ i.i.d. some $F \in \mathcal{F}_c$.

   (b) A *Bayes decision rule* with respect to a prior Λ in a general decision problem with loss function $L(\theta, d)$.

   (c) A *uniformly most powerful unbiased size* α *test* ϕ_0 of $H : \theta \in \Theta_0$ versus $K : \theta \in \Theta_1$.

   (d) A *uniformly most powerful invariant size* α *test* of $H : \theta \in \Theta_0$ versus $K : \theta \in \Theta_1$ in a testing problem with a group G of transformations on the sample space $\mathcal{X}$.

   (e) A maximal invariant for a group of transformations G on a sample space $\mathcal{X}$.

   Solution: See the course notes, Chapters 5 and 6.

2. (36 points) **State** *any three* of the following results:

   (a) Wald's theorem on strong consistency of maximum likelihood estimators.

   (b) The generalized Neyman-Pearson lemma (in the short form stated in the notes).

   (c) A conditional limit theorem about the large sample behavior of posterior distributions. What is the name of this type of theorem?

   (d) A result concerning some optimality property of $\overline{X}_n$ in the context of estimating the mean μ when $X_1, \ldots, X_n$ are i.i.d. $N_d(\mu, \sigma^2 I)$ with σ^2 known and $d \geq 3$

   (e) The Wald-Wolfowitz-Noether-Hájek finite sampling central limit theorem.

   Solution: See the course notes, Chapters 5 and 6.

   Do either problem 3 or problem 4.

3. (40 points) A random variable X takes on the values $1, 2, 3, 4$ with probability distribution $p_0(x)$ or $p_1(x)$ as follows:

x	1	2	3	4
$p_0(x)$	.1	.05	.35	.5
$p_1(x)$	.3	.25	.35	.1

(a) For the usual $0-1$ loss, find a most powerful test of size .10 for testing $H: p = p_0$ versus $K: p = p_1$ and determine its power.

(b) Find a test ϕ which minimizes the sum of risks $E_0\phi + E_1(1-\phi)$. What is the relationship between the minimized sum of risks and the total variation distance between P_0 and P_1 and why does this make intuitive sense?

(c) If the losses are $L(1,1) = L(0,0) = 0$, $L(0,1) = 10$, $L(1,0) = 5$, and the prior is $\lambda = (\lambda_0, \lambda_1) = (.7, .3)$, find the Bayes rule d_B and the minimax rule d_M.

Solution:
(a). The ratios of the probabilities under the two hypotheses are given by the following table

x	1	2	3	4
$p_0(x)$	.1	.05	.35	.5
$p_1(x)$	.3	.25	.35	.1
$p_1(x)/p_0(x)$	3	5	1	1/5

Thus by the Neyman - Pearson lemma the most powerful test of size .10 is $\phi(x) = 1\{x = 2\} + (.5)1\{x = 1\}$. The power of this test is $\beta_\phi = E_1\phi(X) = P_1(X = 2) + .5P_1(X = 1) = .25 + (.5)(.3) = .4$.
(b). The sum of risks $E_0\phi + E_1(1-\phi) = 1 + \sum_x \phi(x)\{p_0(x) - p_1(x)\}$ is minimized by any rule $\phi(x) = 1\{x : p_1(x) > p_0(x)\} = 1_{\{1,2\}}(x)$. The minimum sum of risks is $E_0\phi(X) = 1 + (p_0(1) - p_1(1)) + (p_0(2) - p_1(2)) = 1 - .2 - .2 = .6$. [Note that the total risk of the Neyman-Pearson test of size .10 is $.1 + .6 = .7$.]
(c). For a rule $d = (d_1, d_2, d_3, d_4)$ which chooses 1 with probability d_x when x is observed, the (ordinary) risks are

$$
\begin{aligned}
R(0,d) &= E_0L(0,d) = 10\{d_1(.1) + d_2(.05) + d_3(.35) + d_4(.5) \\
&= \frac{1}{10}\{10d_1 + 5d_2 + 35d_3 + 50d_4\}, \\
R(1,d) &= E_1L(1,d) \\
&= 5\{(1-d_1)(.3) + (1-d_2)(.25) + (1-d_3)(.35) + (1-d_4)(.1)\} \\
&= 5 - 1.5d_1 - 1.25d_2 - 1.75d_3 - .5d_4.
\end{aligned}
$$

Thus the Bayes risk for the prior $\lambda = (.7, .3)$ is

$$\begin{aligned}
\mathcal{R}(\lambda, d) &= .7R(0,d) + .3R(1,d) \\
&= \frac{.7}{10}\{10d_1 + 5d_2 + 35d_3 + 50d_4\} \\
&\qquad + 1.5 - (.3)(1.5d_1 + 1.25d_2 + 1.75d_3 + .5d_4) \\
&= 1.5 + (.7 - .45)d_1 + (.35 - .375)d_2 + (2.45 - .525)d_3 + (3.5 - .15)d_4.
\end{aligned}$$

Since the coefficients of d_1, d_3, and d_4 are positive, while the coefficient of d_2 is negative, the Bayes rule is given by $d = (0, 1, 0, 0)$ with corresponding Bayes risk $\mathcal{R}(\lambda, d_B) = 1.5 - .025 = 1.475$.

To find the minimax rule d_M, we equate the two ordinary risks to find

$$10d_1 + 5d_2 + 35d_3 + 50d_4 = 50 - 15d_1 - 12.5d_2 - 17.5d_3 - 5d_4$$

and solve for d_3 to find that

$$52.5d_3 = 50 - 25d_1 - 17.5d_2 - 55d_4,$$

or equivalently that

$$d_3 = \frac{50}{52.5} - \frac{25}{52.5}d_1 - \frac{17.5}{52.5}d_2 - \frac{55}{52.5}d_4.$$

Substitution into $R(0,d)$ yields a common risk of

$$\begin{aligned}
R(0,d) &= d_1 + .5d_2 + 5d_4 + \frac{35}{10}\left(\frac{50}{52.5} - \frac{25}{52.5}d_1 - \frac{17.5}{52.5}d_2 - \frac{55}{52.5}d_4\right) \\
&= \frac{35 \cdot 5}{52.5} + \left(1 - \frac{35}{10}\frac{25}{52.5}\right)d_1 + \left(.5 - \frac{35}{10}\frac{17.5}{52.5}\right)d_2 + \left(5 - \frac{35}{10}\frac{55}{52.5}\right)d_4 \\
&= \frac{35 \cdot 5}{52.5} + c_1 d_1 + c_2 d_2 + c_4 d_4
\end{aligned}$$

where $c_1 < 0$, $c_2 < 0$ and $c_4 > 0$. Thus the minimax rule has $d_1 = d_2 = 1$, and $d_4 = 0$. This yields

$$d_3 = \frac{50}{52.5} - \frac{15}{52.5} - \frac{17.5}{52.5} = \frac{7.5}{52.5} = \frac{1}{7}.$$

Thus the minimax rule is $d_M = (1, 1, 1/7, 0)$, and the common risk is $R(0,d) = R(1,d) = 1 + .5 + (3.5)(1/7) = 2.0$.

Another way to find the minimax rule is to find a prior distribution $\lambda = (\lambda_0, 1-\lambda_0)$ which is least favorable: Note that with $\phi(x)$ now denoting the probability of choosing 1 when x is observed,

$$R(0,\phi) = 10E_0\phi(X), \qquad R(1,\phi) = 5E_0(1 - \phi(X)),$$

and the Bayes risk is

$$\begin{aligned}\mathcal{R}(\lambda,\phi) &= \lambda_0 10 E_0\phi(X) + (1-\lambda_0)5E_1(1-\phi(X)) \\ &= 5(1-\lambda_0) + \int \phi\{\lambda_0 10 p_0 - 5(1-\lambda_0)p_1\}d\mu,\end{aligned}$$

so the Bayes rules with respect to the prior $(\lambda_0, 1-\lambda_0)$ are of the form

$$\begin{aligned}\phi(x) &= 1\{10\lambda_0 p_0(x) < 5(1-\lambda_0)p_1(x)\} + \gamma 1\{10\lambda_0 p_1(x) = 5(1-\lambda_0)p_0(x)\} \\ &= 1\{p_1(x) > \frac{2\lambda_0}{1-\lambda_0}p_0(x)\} + \gamma 1\{p_1(x) = \frac{2\lambda_0}{1-\lambda_0}p_0(x)\}.\end{aligned}$$

For the choice $\lambda_0 = 1/3$, this becomes

$$\phi(x) = 1\{p_1(x) > p_0(x)\} + \gamma 1\{p_1(x) = p_0(x)\},$$

with ordinary risks

$$\begin{aligned}&R(0,\phi) = 10E_0\phi(X) = 10\{.15 + \gamma(.35)\}, \\ &R(1,\phi) = 5E_1(1-\phi(X)) = 5\{.1 + (1-\gamma).35\}.\end{aligned}$$

Equating these two risks and solving for γ yields $\gamma = 1/7$. For this particular Bayes rule we have $R(0,\phi) = R(1,\phi) = 2 = \mathcal{R}(\lambda,\phi)$, and hence $\lambda = (1/3, 2/3)$ is a least favorable prior and $\phi(x) = 1_{\{1,2\}}(x) + (1/7)1_{\{3\}}(x)$ is minimax.

4. (40 points) Suppose that $X \sim \text{Binomial}(m, p_1)$ and $Y \sim \text{Binomial}(n, p_2)$ are independent. Consider testing $H:\ p_2 \le p_1$ versus $K:\ p_2 > p_1$.

   (a) Write the joint density (probability mass function) $p(x,y;p_1,p_2) \equiv P_{p_1,p_2}(X = x, Y = y)$ in exponential family form $c(\theta,\xi)\exp(\theta U(x,y) + \xi T(x,y))h(x,y)$ where $T(x,y)$ is sufficient for the boundary $\Theta_B = \{(p_1,p_2) \in [0,1]^2 :\ p_1 = p_2\}$ and $U(x,y) = y$.

   (b) Show carefully that testing H versus K is equivalent to testing $H_1 : \theta \le 0$ ($\xi =$ anything) versus $K_1 : \theta > 0$ ($\xi =$ anything).

   (c) What is the conditional distribution of $U(X,Y) = Y$ given $T = t$ under $p_1 = p_2$? Compute it explicitly and give its name.

   (d) Find the UMP unbiased test of size α as explicitly as possible when $m = 3$, $n = 2$, $t = 2$, and $\alpha = 1/10$.

   (e) Relate the conditional distribution of the test statistic involved in (c) to a problem involving sampling without replacement from a finite population. Identify the contents of the urn (i.e. the numbers on the balls in the urn) and calculate the mean and variance of Y given $T = t$ in this conditional distribution.

(f) Use the results of (e) together with the Wald-Wolfowitz-Noether-Hájek CLT to show a conditional (on $T = t$) CLT for Y appropriately centered and normalized if $0 < \liminf(m/N) \le \limsup(m/N) < 1$. [Make sure that you verify the key hypothesis of the theorem.]

Solution: (a) Now

$$
\begin{aligned}
& p(x, y; p_1, p_2) \\
&= \binom{m}{x} p_1^x (1-p_1)^{m-x} \binom{n}{y} p_2^y (1-p_2)^{n-y} \\
&= (1-p_1)^m (1-p_2)^n \exp\left(y \log\left(\frac{p_2/(1-p_2)}{p_1/(1-p_1)} \right) + (x+y)\log(p_1/(1-p_1)) \right) \binom{m}{x}\binom{n}{y} \\
&= c(\theta, \xi) \exp(\theta U(x,y) + \xi T(x,y)) h(x,y)
\end{aligned}
$$

where

$$
\begin{aligned}
U(x,y) &= y, & \theta &= \theta(p_1, p_2) = \log\left(\tfrac{p_2/(1-p_2)}{p_1/(1-p_1)} \right) \\
T(x,y) &= x + y, & \xi &= \xi(p_1, p_2) = \log(p_1/(1-p_1)),
\end{aligned}
$$

(b) Since $\theta = 0$ on the boundary set $p_1 = p_2$, and since $\theta(p_1, p_2)$ is a monotone increasing function of p_2 for each fixed p_1, it follows that $H: \ p_2 \le p_1$ versus $K: \ p_2 > p_1$ corresponds to $H_1 : \theta \le 0$ versus $K_1 : \theta > 0$.
(c) The conditional distribution of $U = Y$ given $T = X + Y$ under $p_1 = p_2$ is given by

$$
\begin{aligned}
p(y|t) &= P_{(p_1,p_1)}(Y = y | T = t) = \frac{P_{(p_1,p_1)}(Y = y, T = X + Y = t)}{P_{(p_1,p_1)}(T = t)} \\
&= \frac{P_{(p_1,p_1)}(Y = y, X = t - y)}{P_{(p_1,p_1)}(T = t)} = \frac{\binom{m}{t-y} p_1^{t-y} (1-p_1)^{m-(t-y)} \binom{n}{y} p_1^y (1-p_1)^{n-y}}{\binom{N}{t} p_1^t (1-p_1)^{N-t}} \\
&= \frac{\binom{n}{y}\binom{m}{t-y}}{\binom{N}{t}}, \qquad \text{for } \ 0 \vee (t-m) \le y \le t \wedge n
\end{aligned}
$$

where $N \equiv m+n$. This is the hypergeometric distribution with parameters N, n, t (total number of balls N, number of "white balls" n (and number of black balls $m = N - n$), and number of draws t), describing the distribution of the total number of white balls drawn in a sample of size t drawn without replacement from an urn containing n white balls and $m = N - n$ black balls.
(d) The UMP unbiased test of H versus K rejects when $Y > c_t$ where c_t is determined from the hypergeometric distribution in (c) so that $P(Y > c_t | T =$

$t) = \alpha$ (or a randomized version of this) When $m = 3$, $n = 2$, $t = 2$, and $\alpha = .1$, the hypergeometric distribution become

$$P(Y = 2|T = 2) = \frac{\binom{2}{2}\binom{3}{0}}{\binom{5}{2}} = \frac{1}{10},$$
$$P(Y = 1|T = 2) = \frac{\binom{2}{1}\binom{3}{1}}{\binom{5}{2}} = \frac{6}{10},$$
$$P(Y = 0|T = 2) = \frac{\binom{2}{0}\binom{3}{2}}{\binom{5}{2}} = \frac{3}{10},$$

so $c_t = 1$ and rejecting when $Y > 1$ (i.e. when $Y = 2$) gives conditional level $\alpha = .1$.

(e) For the description, see the solution of (c) above; we can also identify the "white balls" as those with the number 1, and the "black balls" as those with the number zero. Then $E(Y|T = t) = t(n/N)$ while $Var(Y|T = t) = (1 - \frac{t-1}{N-1})t\sigma_z^2$ where

$$\begin{aligned}
\sigma_z^2 &= \frac{1}{N}\sum_1^N (z_i - \overline{z})^2 = \frac{1}{N}\{n(1 - n/N)^2 + m(0 - n/N)^2\} \\
&= (n/N)(1 - n/N)^2 + (m/N)(0 - n/N)^2 \\
&= (n/N)(1 - n/N)\{(1 - n/N) + n/N\} = (n/N)(1 - (n/N)).
\end{aligned}$$

(f) Since we can view the conditional distribution of Y given $T = t$ as the total in drawing a sample of size t balls without replacement from an urn with balls labeled with z_i's, n of which are 1's and m of which are 0's, and by the calculation of the conditional mean and variance in (e),

$$\sigma_N^2 = Var[(Y/T)|T = t] = \left(1 - \frac{t-1}{N-1}\right)\frac{\sigma_z^2}{t} = \left(1 - \frac{t-1}{N-1}\right)\frac{(n/N)(1 - n/N)}{t}$$

it follows from the W-W-N-H finite sampling CLT that with $\bar{Y} = Y/t$

$$\frac{\bar{Y} - \mu_z}{\sigma_N} = \frac{Y/t - (n/N)}{\sigma_N} \to_d N(0, 1)$$

as $t \to \infty$ and $N \to \infty$ if $0 < \liminf((N - t)/N) \le \limsup((N - t)/N) < 1$ and the Noether condition holds. But the latter is easily satisfied in this case since

$$\begin{aligned}
\eta_N = \frac{\max_{i \le N}|z_i - \overline{z}|^2}{\sum_1^N (z_i - \overline{z})^2} &= \frac{|1 - (n/N)|^2 \vee |0 - (n/N)|^2}{n(1 - (n/N))^2 + m(0 - n/N)^2} \\
&= \frac{(m/N) \vee (n/N)}{m \wedge n} \le \frac{1}{m \wedge n} \to 0
\end{aligned}$$

as $m \wedge n \to \infty$. Note that

$$0 < \liminf((N-T)/N) \leq \limsup((N-T)/N) < 1 \tag{1}$$

holds almost surely if and only if

$$0 < \liminf(T/N) \leq \limsup(T/N) < 1, \tag{2}$$

and since $T/N = (X+Y)/N = (m/N)(X/m) + (n/N)(Y/n)$ where $X/m \to_{a.s.} p_1$ and $Y/n \to_{a.s.} p_2$, (1) and (2) hold if $0 < p_1 \vee p_2$ and $p_1 \wedge p_2 < 1$ and $0 < \liminf(n/N) \leq \limsup(n/N) < 1$.

Do either problem 5 or problem 6.

5. (40 points).

   (a) State the short form of the generalized NP lemma.

   (b) Prove the generalized NP lemma you stated in (a).

   (c) A test ϕ_0 is said to be a locally best unbiased size α test of $H : \theta = \theta_0$ versus $K : \theta \neq \theta_0$ if among all tests ϕ satisfying

$$\beta_\phi(\theta_0) = \alpha \quad \text{and} \quad \beta'_\phi(\theta_0) = 0, \tag{3}$$

   the test ϕ_0 maximizes $\beta''_\phi(\theta_0)$.; i.e. $\beta''_{\phi_0}(\theta_0) \geq \beta''_\phi(\theta_0)$ for all tests ϕ satisfying both equalities in (3). Assuming that the power function $\beta_\phi(\theta)$ of any test ϕ admits two continuous derivatives which may be passed under the integral sign (so $\beta'_\phi(\theta) = \int \phi(x)(\partial/\partial\theta)f(x|\theta)dx$ and $\beta''_\phi(\theta) = \int \phi(x)(\partial^2/\partial\theta^2)f(x|\theta)dx$, use the generalized NP lemma to derive the form of the locally best unbiased test of $H : \theta = \theta_0$ versus $K : \theta \neq \theta_0$.

Solution: (a) and (b): See the course notes.
(c) We want to maximize

$$\beta''_\phi(\theta_0) = \int \phi(x) \frac{\partial^2}{\partial\theta^2} f(x|\theta)\big|_{\theta=\theta_0} dx$$

subject to

$$\beta'_\phi(\theta_0) = \int \phi(x) \frac{\partial}{\partial\theta} f(x|\theta)\big|_{\theta=\theta_0} dx = 0$$

(to achieve local unbiasedness), and

$$\beta_\phi(\theta_0) = \int \phi(x) f(x|\theta_0) dx = \alpha$$

to achieve the right size (namely α) at $\theta = \theta_0$. Thus we can apply the generalized Neyman-Pearson lemma with

$$\begin{aligned} f_0(x) &\equiv \frac{\partial^2}{\partial\theta^2} f(x|\theta)\big|_{\theta=\theta_0}, \\ f_1(x) &\equiv \frac{\partial}{\partial\theta} f(x|\theta)\big|_{\theta=\theta_0}, \quad \text{and} \\ f_2(x) &\equiv f(x|\theta_0). \end{aligned}$$

Thus the locally best unbiased test at $\theta = \theta_0$ is of the form

$$\phi_0(x) = \begin{cases} 1, & \text{if } f_0(x) > k_1 f_1(x) + k_2 f_2(x), \\ \gamma(x), & \text{if } f_0(x) = k_1 f_1(x) + k_2 f_2(x), \\ 0, & \text{if } f_0(x) < k_1 f_1(x) + k_2 f_2(x), \end{cases}$$

where we choose k_1, k_2, and $\gamma(x)$ to ensure $\beta'_\phi(\theta_0) = 0$ and $\beta_\phi(\theta_0) = \alpha$. It is instructive to re-express this test in terms of the derivatives of $\log f(x|\theta)$. See e.g. Ferguson (Math. Statist.), pages 238 - 240 for more on this and the choice of k_1 and k_2.

6. (40 points) Suppose that X is a random variable with density $p(\cdot;\theta)$ given by

$$p(x;\theta) = \frac{1}{\pi}\frac{1}{\sqrt{x(1-x)}} \cdot \frac{\theta}{1-(1-\theta^2)x} 1_{(0,1)}(x), \quad \theta \in (0,\infty). \tag{4}$$

This is the distribution of the total time spent in $(0,\infty)$ up to time $t = 1$ by a skew Brownian motion process with skewing parameter $\theta = \sigma_+/\sigma_-$ where σ_+^2 is the variance parameter for the positive space axis and σ_-^2 is the variance parameter for the negative space axis. Note that for $\theta = 1$,

$$p(x,1) = \frac{1}{\pi}\frac{1}{\sqrt{x(1-x)}} 1_{(0,1)}(x)$$

is the Beta$(1/2, 1/2)$ density corresponding to the arcsin distribution

$$F_1(x) = P_1(X \le x) = \frac{2}{\pi}\arcsin(\sqrt{x}).$$

The mean and variance of X with density $p(\cdot;\theta)$ are $E_\theta X = 1/(1+\theta)$ and $Var_\theta(X) = 2^{-1}\frac{1}{1+\theta} \cdot \frac{\theta}{1+\theta}$.

(a) Show that the family $\mathcal{P} = \{p(\cdot;\theta) : \theta \in (0,\infty)\}$ has monotone likelihood ratio in $T(x) = 1/x$. [Hint: rewrite the last factor in (4) in terms of $1/x$.]

(b) Find the UMP size .05 test of $H: \theta \le 1 \equiv \theta_0$ versus $K: \theta > 1$? Specify your test completely, including the constant(s).

(c) Compute the power function of the UMP test in (b) as explicitly as possible.

(d) Now suppose that $X_1, \ldots, X_n$ are i.i.d. with density $p(\cdot;\theta)$. Find the form of the locally most powerful test of $H: \theta \le 1$ versus $K: \theta > 1$ based on $X_1, \ldots, X_n$, and use the central limit theorem to find appropriate constants so that your test has approximate size $\alpha = .05$.

Solution: (a) Suppose that $0 < \theta < \theta' < \infty$. Then

$$\begin{aligned} \frac{p(x;\theta')}{p(x;\theta)} &= \frac{\theta'}{1-(1-\theta'^2)x} \cdot \frac{1-(1-\theta^2)x}{\theta} = \frac{\theta'}{\theta} \cdot \frac{(1/x)-(1-\theta^2)}{(1/x)-(1-\theta'^2)} \\ &\equiv \frac{\theta'}{\theta} \cdot \frac{t-(1-\theta^2)}{t-(1-\theta'^2)}. \end{aligned}$$

Thus

$$\begin{aligned}
\frac{d}{dt}\log\left(\frac{p(1/t;\theta')}{p(1/t;\theta)}\right) &= \frac{1}{t-(1-\theta^2)} - \frac{1}{t-(1-\theta'^2)} \\
&= \frac{\theta'^2-\theta^2}{(t-(1-\theta^2))(t-(1-\theta^2))} \\
&> 0.
\end{aligned}$$

Thus the family $\mathcal{P} = \{p(\cdot;\theta) : \theta > 0\}$ has MLE in $T(x) = 1/x$.
(b) It follows from the Karlin-Rubin theorem that the UMP size α test of $H : \theta \le 1$ versus $K : \theta > 1$ is of the form $\phi(X) = 1\{1/X > c\} + \gamma 1\{1/X = c\} = 1\{X < 1/c\} + \gamma 1\{X = 1/c\}$ where c and γ are determined by $\alpha = E_{\theta_0=1}\phi(X) = P_1(X < 1/c) + \gamma P_1(X = 1/c)$. Since the distribution of X is continuous we can take $\gamma = 0$ and choose $1/c \equiv \tilde{c}$ to satisfy $P_1(X < \tilde{c}) = \alpha = .05$. But since we know that $P_1(X \le x) = (2/\pi)\arcsin(\sqrt{x})$, this yields $\tilde{c} = \tilde{c}_\alpha = (\sin(\pi\alpha/2))^2 = (\sin(\pi/40))^2$.
(c) The power function of the test in (b) is just

$$\begin{aligned}
\beta_\phi(\theta) &= E_\theta\phi(X) = P_\theta(X < \tilde{c}_\alpha) = \int_0^{\tilde{c}_\alpha} p(x;\theta)dx \\
&= \int_0^{\tilde{c}_\alpha} \frac{1}{\pi}\frac{1}{\sqrt{x(1-x)}} \cdot \frac{\theta}{1-(1-\theta^2)x}dx \\
&= 1 - \frac{2}{\pi}\arctan\left(\frac{\cot(\alpha\pi/2)}{\theta}\right)
\end{aligned}$$

after three changes of variables: $x = t^2$; followed by $t = 1/y$; followed by $v = \sqrt{y^2-1}$. See Figure 1 for a plot of this power function.

Here is a detailed derivation of the power formula: by letting $x = t^2$ we find that

$$
\begin{aligned}
&\int_0^{\tilde{c}_\alpha} \frac{1}{\pi}\frac{1}{\sqrt{x(1-x)}} \cdot \frac{\theta}{1-(1-\theta^2)x}dx \\
&= \int_0^{\sin(\alpha\pi/2} \frac{1}{\pi}\frac{\theta 2t}{\sqrt{t^2-t^4}\cdot(1-(1-\theta^2)t^2)}dt \\
&= \int_0^{\sin(\alpha\pi/2} \frac{2}{\pi}\frac{\theta}{\sqrt{1-t^2}\cdot(1-(1-\theta^2)t^2}dt \\
&= \frac{2\theta}{\pi}\int_0^{\sin(\alpha\pi/2)} \frac{1}{t^2\sqrt{1-t^2}\cdot(1/t^2-(1-\theta^2))}dt \\
&= \frac{2\theta}{\pi}\int_0^{1/\sin(\alpha\pi/2)} \frac{-1/y^2}{(1/y^2)\sqrt{1-1/y^2}\cdot(y^2-(1-\theta^2)}dy, \quad t = 1/y, \\
&= \frac{2\theta}{\pi}\int_{1/\sin(\alpha\pi/2)}^{\infty} \frac{y}{\sqrt{y^2-1}\cdot(y^2-1+\theta^2)}dy \\
&= \frac{2\theta}{\pi}\int_{\cot(\alpha\pi/2)}^{\infty} \frac{1}{v^2+\theta^2}dv, \quad v = \sqrt{y^2-1} \\
&= \frac{2\theta}{\pi}\frac{1}{\theta}\arctan(v/\theta)\Big|_{\cot(\alpha\pi/2)}^{\infty} \\
&= \frac{2}{\pi}\left(\frac{\pi}{2} - \arctan\left(\frac{\cot(\alpha\pi/2)}{\theta}\right)\right) \\
&= 1 - \frac{2}{\pi}\arctan\left(\frac{\cot(\alpha\pi/2)}{\theta}\right).
\end{aligned}
$$

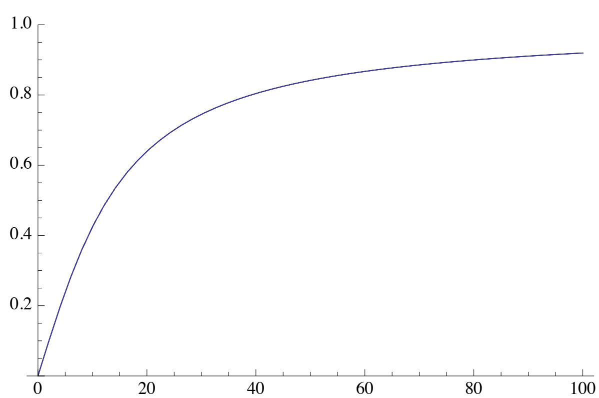

Figure 1: Plot of power, UMP test for $n = 1$.

(d) As we have seen in the context of the Cauchy location family, the locally most powerful test of $H : \theta = \theta_0 = 1$ versus $K : \theta > 1$ is of the form "reject H

if $\dot{\mathbf{l}}_{n\theta}(\underline{X};\theta_0) > k$" where $\dot{\mathbf{l}}_{n,\theta}$ is the score function based on all the data and k is a constant to be determined so that the test has (at least approximately) size α. In this present case, the score for one observation is

$$\begin{aligned}
\dot{\mathbf{l}}_\theta(x;\theta) &= \frac{1}{\theta} - \frac{2\theta x}{1-(1-\theta^2)x} = \frac{1-(1+2\theta-\theta^2)x}{\theta(1-(1-\theta^2)x} \\
&= 1-2x \qquad \text{when} \ \ \theta = \theta_0 = 1.
\end{aligned}$$

Thus the locally most powerful test of $H : \theta = 1$ versus $K : \theta > 1$ is of the form "reject $H : \theta = 1$ if $\sum_1^n (1-2X_i) > k$; or, equivalently when $\overline{X}_n - (1/2) < k'$; or, equivalently, when $Z_n \equiv \sqrt{n}(\overline{X}_n - 1/2)/\sqrt{1/8} < k''$. Under $\theta_0 = 1$ the X_i's have $E_1(X_1) = 1/2$ and $Var_1(X_1) = 1/8$, and therefore the CLT implies that the test

$$\phi(\underline{X}) = 1\{Z_n < -z_\alpha\}, \quad \text{where } z_\alpha \equiv \Phi^{-1}(1-\alpha)$$

has $E_1\phi(\underline{X}) \to \alpha$.

Do either problem 7 or problem 8.

7. (40 points) Suppose that an urn contains N balls with the numbers $a_N(1), \ldots, a_N(N)$ written on the balls. Suppose that a sample of n balls is drawn from the urn without replacement: let the numbers on the sampled balls be $Y_1, \ldots, Y_n$, and let $\overline{Y}_n = n^{-1}\sum_{i=1}^n Y_i$.
(a) What is the mean of $\overline{Y}_n$?
(b) What is the variance of $\overline{Y}_n$?
(c) If $\underline{R} = (R_1, \ldots, R_N)$ is a random permutation of $\{1, \ldots, N\}$, what is the relationship between $(Y_1, \ldots, Y_n)$ and $(a_N(R_1), \ldots, a_N(R_n))$?
(d) Under some condition on the numbers $\{a_N(i)\}$, a CLT holds for an appropriately standardized version of $\overline{Y}_n$, and hence also for $\sum_{j=1}^n a_N(R_j)$. State this condition and the theorem.
(e) If $a_N(j) = j$, for $j = 1, \ldots, N$, compute the mean and variance in (a) and (b), and make the conclusion of the CLT in (d) explicit.]
[Hint: recall that $\sum_{j=1}^N j = N(N+1)/2$ and $\sum_{j=1}^N j^2 = N(N+1)(2N+1)/6$.]

Solution: (a) $E(\overline{Y}_n) = \overline{a}_N \equiv N^{-1}\sum_{i=1}^N a_N(i)$.
(b) $Var(\overline{Y}_n) = n^{-1}\sigma_a^2(1 - \frac{n-1}{N-1}) \equiv \sigma_N^2$ where $\sigma_a^2 = N^{-1}\sum_{i=1}^N (a_N(i) - \overline{a}_N)^2$.
(c) The random vectors $\underline{Y} = (Y_1, \ldots, Y_n)$ and $(a_N(R_1), \ldots, a_N(R_n))$ satisfy $(Y_1, \ldots, Y_n) \stackrel{d}{=} (a_N(R_1), \ldots, a_N(R_n))$.
(d) If $0 < \liminf(n/N) \le \limsup(n/N) < 1$, then with σ_N^2 as in (b)

$$\frac{\overline{Y}_n - \overline{a}_N}{\sigma_N} \to_d Z \sim N(0,1)$$

if and only if

$$\eta_N \equiv \frac{\max_{1\le i\le N}(a_N(i) - \overline{a}_N)^2}{\sum_{i=1}^N (a_N(i) - \overline{a}_N)^2} \to 0. \tag{5}$$

(e) When $a_N(j) = j$ for $j \in \{1, \ldots, N\}$ we have $\overline{a}_N = (N+1)/2$, and $N^{-1}\sum_{i=1}^N a_N(i)^2 = (N+1)(2N+1)/6$, so (after a bit of algebra) $\sigma_a^2 = (N^2-1)/12$, and the Noether condition (5) holds since

$$\eta_N = \frac{(N-1)^2/4}{N\sigma_a^2} = \frac{(N-1)^2/4}{N(N^2-1)/12} \to 0$$

as $N \to \infty$. In this case $\sum_{i=1}^N Y_i \stackrel{d}{=} \sum_{i=1}^n a_N(R_i)$ where $P(\underline{R} = \underline{r}) = 1/N!$ for each permutation $\underline{r}$ of $\{1, \ldots, N\}$.

8. (40 points) Suppose that $X_i \sim \text{Poisson}(\mu i)$, $i = 1, \ldots, m$ are independent, and that $Y_j \sim \text{Poisson}(\nu j)$, $j = 1, \ldots, n$ are also independent and independent of the X_i's. Consider testing $H : \mu \geq \nu$ versus $K : \mu < \nu$.

   (a) Show that we can reduce by sufficiency to $R \equiv \sum_{i=1}^m X_i$ and $S \equiv \sum_{j=1}^n Y_j$.

   (b) What are the distributions of R and S?

   (c) Find a level α UMP - unbiased test of H versus K and indicate exactly how to carry it out, identifying the relevant conditional distribution explicitly.

   (d) Describe the Bayes test of H versus K (for $0 - 1$ loss assuming a prior distribution Λ of (μ, ν). Can you describe the rejection region of this test when Λ is the product of two independent Γ priors (i.e. $\mu \sim \Gamma(\alpha, \beta)$ and $\nu \sim \Gamma(\gamma, \delta)$)?

Solution: (a) Note that the joint mass function of the X_i's is given by

$$
\begin{aligned}
p_\mu(\underline{x}) &= \prod_{i=1}^m \exp(-i\mu) \frac{(i\mu)^{x_i}}{x_i!} = \prod_{i=1}^m \exp(-i\mu) \exp(x_i \log(i\mu)) \frac{1}{x_i!} \\
&= \exp(-\mu \sum_{i=1}^m i + \sum_{i=1}^m x_i \log \mu + \sum_{i=1}^m x_i \log i) \frac{1}{\prod_{i=1}^m x_i!} \\
&= c(\mu) \exp(\sum_1^m x_i (\log \mu)) h(\underline{x}),
\end{aligned}
$$

and hence $R = \sum_{i=1}^m X_i$ is sufficient for μ. Similarly, $S = \sum_{j=1}^n Y_j$ is sufficient for ν.

(b) Since sums of independent Poisson random variables are again poisson with intensity parameter the sum of the intensity parameters of the summands, we see that

$$
R = \sum_{i=1}^m X_i \sim \text{Poisson}(\mu \sum_1^m i) = \text{Poisson}(\mu m(m+1)/2) = \text{Poisson}(a\mu)
$$

where $a \equiv m(m+1)/2$. Similarly, $S = \sum_{j=1}^n Y_j \sim \text{Poisson}(b\nu)$ where $b \equiv n(n+1)/2$.

(c) The joint mass function of R and S is

$$\begin{aligned}
p(r, s; \mu, \nu) &= e^{-a\mu}\frac{(a\mu)^r}{r!}e^{-b\nu}\frac{(b\nu)^s}{s!} \\
&= c(\mu, \nu)\exp(r\log(a\mu) + s\log(b\nu))h(r, s) \\
&= c(\mu, \nu)\exp(r\log(\mu) + s\log(\nu))\tilde{h}(r, s) \\
&= c(\mu, \nu)\exp(s(\log\nu - \log\mu) + s\log\mu + r\log\mu)\tilde{h}(r, s) \\
&= c(\mu, \nu)\exp(\theta s + \eta(r + s))\tilde{h}(r, s)
\end{aligned}$$

where $\theta \equiv \log(\nu/\mu)$, $\eta \equiv \log\mu$, and we see that $T \equiv R + S$ is sufficient for $\Theta_B \equiv \overline{\Theta}_0 \cap \overline{\Theta}_1$, and testing H versus K is equivalent to testing $\theta \le 0$ versus $\theta > 0$. Thus the UMP unbiased test of H versus K is of the form

$$\phi(S) = 1\{S > c_T\} + \gamma_T 1\{S = c_T\}$$

where c_T and γ_T are determined so that $E\{\phi(S)|T\} = \alpha$. But

$$(S|T) \sim \text{Binomial}\left(T, \frac{\nu b}{\mu a + \nu b}\right) = \text{Binomial}\left(T, \frac{b}{a + b}\right)$$

under $\mu = \nu$.

(d) When (μ, ν) have the prior distribution Λ, the Bayes test of H versus K is "reject H if $P(\nu > \mu|R, S) > P(\nu \le \mu|R, S)$", or, equivalently, if $P(\nu > \mu|R, S) > 1/2$. When Λ has density $\lambda(\mu, \nu)$ given by the product of two Gamma densities as described, then

$$\begin{aligned}
p_\mu(r)\lambda_{\alpha,\beta}(\mu) &= \exp(-a\mu)\frac{(a\mu)^r}{r!} \cdot \frac{\beta(\beta\mu)^{\alpha-1}}{\Gamma(\alpha)}e^{-\beta\mu} \\
&\propto \mu^{\alpha+r-1}\exp(-(a + \beta)\mu),
\end{aligned}$$

so we see that the posterior density of μ given $R = r$ is Gamma$(\alpha + r, a + \beta)$. Similarly, the posterior density of ν given $S = s$ is Gamma$(\gamma + s, b + \delta)$. Thus we find that the posterior probability in question is, with $a = m(m + 1)/2$, $b = n(n + 1)/2$,

$$P(\nu > \mu|R, S) = P(\text{Gamma}(\gamma + S, b + \delta) > \text{Gamma}(\alpha + R, a + \beta))$$

where the two Gamma random variables are independent.