Master's Exam - Spring 2012
March 22, 2012
1:00 pm - 5:00 pm

Answers are sketched.

Name:______________________

A. The number of points for each problem is given.

B. There are problems with varying numbers of parts.

Problems 1 - 6 Probability (60 points)
Problems 7 - 12 Statistics (60 points)

C. Write your answers on the exam paper itself. If you need more room you may use the extra sheets provided. Answer as many questions as you can on each part. Tables are provided.

Good Luck!

1. (6 pts.) X and Y are independent random variables with X distributed Poisson with mean λ_1 and Y distributed Poisson with mean λ_2. Prove that $X + Y$ is distributed Poisson with mean $\lambda_1 + \lambda_2$.

$$P(X+Y=z) = \sum_{x=0}^{z} \frac{\lambda_1^x e^{-\lambda_1}}{x!} \frac{\lambda_2^{z-x} e^{-\lambda_2}}{(z-x)!}$$

$$= \frac{\lambda_2^z e^{-(\lambda_1+\lambda_2)}}{z!} \sum_{x=0}^{z} \left(\frac{\lambda_1}{\lambda_2}\right)^x \frac{z!}{x!\,(z-x)!}$$

$$= \frac{\lambda_2^z e^{-(\lambda_1+\lambda_2)}}{z!} \left(1 + \frac{\lambda_1}{\lambda_2}\right)^z \quad \text{by Binomial Thm}$$

$$= \frac{(\lambda_1+\lambda_2)^z e^{-(\lambda_1+\lambda_2)}}{z!}, \quad z = 0, 1, 2, \ldots$$

which is seen to be the Poisson pmf with mean $\lambda_1 + \lambda_2$.

2. Consider independent tosses of a two-sided coin where on a single toss the probability of Heads (H) is 0.6 and probability of Tails (T) is 0.4.

(a) (4 pts.) What is the probability of more Heads than Tails in five tosses?

X = No. of H's $\qquad X \sim B(5, .6)$

$$P(X \geq 3) = \sum_{x=3}^{5} \binom{5}{x} .6^{x} .4^{5-x} = 0.683$$

(b) (6 pts.) The coin is given independent tosses until Heads is observed on two consecutive tosses or Tails is observed on two consecutive tosses. What is the probability that it takes more than six tosses to observe either two Heads in a row or two Tails in a row?

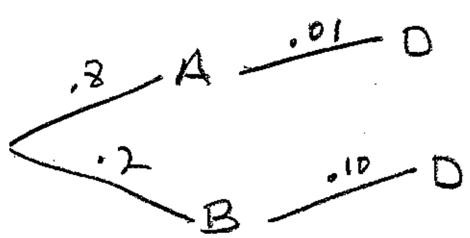
N = No of tosses until either 2 H's in a row or 2 T's in a row.

$[N > 6] = \{HTHTHT, THTHTH\}$ so

$$P[N > 6] = 2 \times .6^3 \times .4^3 = .0276$$

3. Some research shows that people with different genotypes A and B could have distinct infection rates for a disease. Assume that the probability that a person with genotype A is infected by the disease is 0.01 while the probability is 0.1 for people who have genotype B. Suppose that 80% people in the population are of genotype A and 20% are of genotype B.

(a) (3 pts.) What is the probability that a randomly chosen person is infected by the disease?



$$P(D) = (.8)(.01) + (.2)(.10) = .028$$

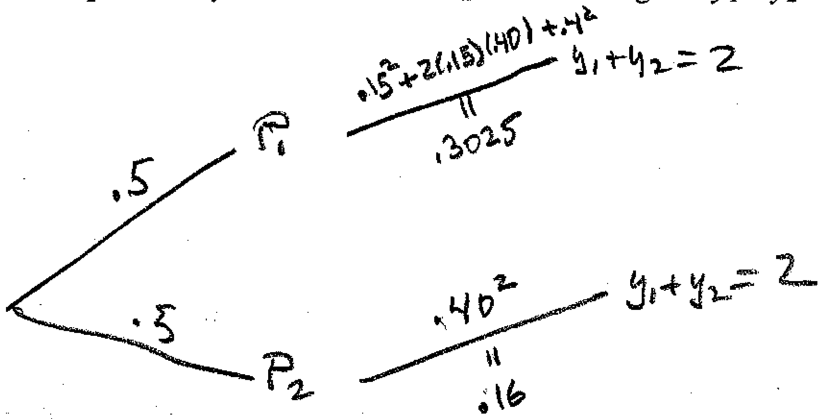
(b) (3 pts.) If a person is infected by the disease, what is the conditional probability that this person having genotype A?

$$P(A|D) = \frac{.008}{.028} = \frac{2}{7}$$

4. (10 pts.) Here are two probability distributions P_1 and P_2 on the sample space $\{(0, 0), (0, 1), (1, 0), (1, 1)\}$.

(x, y)	(0, 0)	(0, 1)	(1, 0)	(1, 1)
P_1	.20	.15	.25	.40
P_2	.40	.40	.20	.00

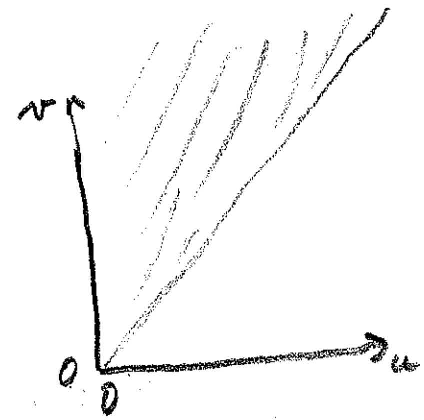
A distribution is selected from $\{P_1, P_2\}$ with equal probability 1/2 on each distribution. Conditional on the choice of distribution, two independent observations (x_1, y_1), (x_2, y_2) are generated. Calculate the probability that distribution P_2 was chosen given $y_1 + y_2 = 2$.



$$P(y_1 + y_2 = 2) = .5(.3025) + .5(.16)$$
$$= .23125$$

$$P(P_2 \mid y_1 + y_2 = 2) = \frac{(.5)(.16)}{.23125} = .346$$

5. The random vector (U, V) has the probability density function



$$g(u,v) = 2e^{-(u+v)},\ 0 \le u \le v < \infty$$
$$= 0, \text{ otherwise.}$$

(a) (7 pts.) Find the probability density of U.

$$g_U(u) = \int_u^{\infty} 2e^{-(u+v)}\,dv$$

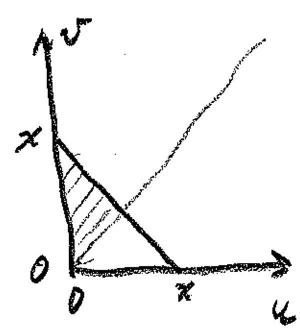
$$= 2e^{-u}\int_u^{\infty} e^{-v}\,dv$$

$$= 2e^{-u}\left[-e^{-v}\right]_u^{\infty}$$

$$= 2e^{-u}\cdot e^{-u}$$

$$= 2e^{-2u}\qquad, u > 0$$

(b) (7 pts.) Find the probability density of $U + V$.



$$P[U+V \le x] = \int_0^{x/2}\int_u^{x-u} 2e^{-(u+v)}\,dv\,du$$

$$= 2\int_0^{x/2} e^{-u}\left[-e^{-v}\right]_u^{x-u} du$$

$$= 2\int_0^{x/2}\left[e^{-2u} - e^{-x}\right] du$$

$$= -e^{-2u}\Big]_0^{x/2} - 2ue^{-x}\Big]_0^{x/2}$$

$$= 1 - e^{-x} - xe^{-x}\quad, x > 0$$

So pdf is

$$\frac{d}{dx}\left(1-e^{-x}-xe^{-x}\right)$$
$$= xe^{-x},\ x > 0$$

6. Let $\{X_n\}$, $n = 1, 2, \ldots$ be a sequence of independent exponentially distributed random variables where X_n has mean n, $n = 1, 2, \ldots$ Thus, the cumulative distribution function of X_n is $F_n(x) = 1 - e^{-x/n}$, $x > 0$. Define $Z_n := \min\{X_1, X_2, \ldots, X_n\}, n = 1, 2, \ldots$

(a) (7 pts.) Determine the probability density function of Z_n.

$$P(Z_n > z) = P(Z_1 > z, Z_2 > z, \ldots, Z_n > z)$$
$$= e^{-z/1} \cdot e^{-z/2} \cdots e^{-z/n} = e^{-(1+\frac{1}{2}+\cdots+\frac{1}{n})z}$$

so Z_n has pdf $f(z) = \frac{d}{dz}\left[1 - e^{-(1+\frac{1}{2}+\cdots+\frac{1}{n})z}\right]$

$$= \left(\sum_1^n \frac{1}{j}\right) e^{-\left(\sum_1^n \frac{1}{j}\right)z}, \quad z > 0$$

(b) (7 pts.) Prove that $Z_n \to 0$ in probability as $n \to \infty$.

$\epsilon > 0$

$$P(|Z_n - 0| > \epsilon) = P(Z_n > \epsilon) = e^{-\left(\sum_1^n \frac{1}{j}\right)\epsilon}$$

The harmonic series $\sum_1^n \frac{1}{j} \to \infty$ as $n \to \infty$

so that $P(|Z_n - 0| > \epsilon) \to 0$ as $n \to \infty$

7. (6 pts.) A nationwide poll of 625 students showed that 400 believed that education was very important in their future, 150 believed that it was moderately important and the rest thought that it was unimportant. Consider this to be a random sample. Find a 90 % confidence interval estimate of p, the proportion of all students nationwide who believed that education was very important in their future.

$\hat{p} = \frac{400}{625} = .64$ Using normal approx to the Binomial, an approx 90% CI estmate of p is

$$.64 \pm 1.645 \sqrt{\frac{(.64)(.36)}{625}}$$

$$.64 \pm .032$$

8. Let $X_1, X_2, \ldots, X_n$ be independent and identically distributed (*iid*) random variables with common *pdf*

$$f(x; \mu, \lambda) = \frac{1}{\lambda} \exp\left(\frac{-(x-\mu)}{\lambda}\right), \quad x > \mu \quad \text{where } \lambda > 0 \text{ and } -\infty < \mu < \infty.$$

(a) (5 pts.) What is the log-likelihood function for μ and λ?

The likelihood function is

$$L(\mu, \lambda) = \prod_{i=1}^{n} \frac{1}{\lambda} \exp\left(-\frac{x_i - \mu}{\lambda}\right) I(x_i \geq \mu) = \left(\frac{1}{\lambda}\right)^n \exp\left(-\sum_{i=1}^{n} \frac{x_i - \mu}{\lambda}\right) I(\min_i x_i > \mu)$$

So the log-likelihood function is

$$\ell(\mu, \lambda) = -n \log \lambda - \frac{1}{\lambda} \sum_{i=1}^{n} (x_i - \mu) \quad \text{for } \mu < \min_i x_i$$

$$= -\infty \quad \text{otherwise}$$

8. (cont).

(b) (8 pts.) Find the maximum likelihood estimates (MLE) for μ and λ?

Fix λ, the likelihood function $L(\mu,\lambda)$ is an increasing function of μ.

Therefore, the MLE for μ is $\hat{\mu} = \min_i X_i$. Because for any λ, the likelihood is maximized at $\hat{\mu}$, the MLE for μ is $\hat{\mu}$.

Fix $\mu = \hat{\mu}$, the MLE for λ satisfies

$$\frac{\partial \ell(\hat{\mu},\lambda)}{\partial \lambda} = -\frac{n}{\lambda} + \frac{1}{\lambda^2}\sum_{i=1}^{n}(X_i - \hat{\mu}) = 0$$

So $\hat{\lambda} = \frac{1}{n}\sum_{i=1}^{n}(X_i - \hat{\mu})$.

$$\frac{\partial^2 \ell(\hat{\mu},\lambda)}{\partial \lambda} = \frac{n}{\lambda^2} - \frac{2n}{\lambda^3}\frac{1}{n}\sum_{i=1}^{n}(X_i - \hat{\mu})$$

$$\left.\frac{\partial^2 \ell(\hat{\mu},\lambda)}{\partial \lambda}\right|_{\lambda=\hat{\lambda}} = \frac{n}{\hat{\lambda}^2} - \frac{2n}{\hat{\lambda}^2} = -\frac{n}{\hat{\lambda}^2} < 0 .$$

Therefore, the MLEs for (μ,λ) are $(\hat{\mu},\hat{\lambda})$.

8. (cont).

(c) (8 pts.) Assume $\mu = 0$. Construct a 90% confidence interval for λ based on $\{X_1, X_2, \ldots, X_n\}$.

If $\mu=0$, X_i has a exponential(λ). Since λ is a scale parameter, a pivotal quantity is

$$Q(X_1, \cdots, X_n; \lambda) = \sum_{i=1}^{n} \frac{X_i}{\lambda} .$$

It's easy to know $Q(X_1, \cdots X_n; \lambda)$ has a Gamma$(n, 1)$ distribution. Let $T_{n,1;\alpha/2}$ and $T_{n,1;1-\alpha/2}$ be the $\alpha/2$ and $1-\alpha/2$ quantiles of a Gamma$(n, 1)$ distribution. Then a 90% confidence interval for λ is

$$\left[\frac{\sum_{i=1}^{n} X_i}{T_{n,1;0.95}}, \frac{\sum_{i=1}^{n} X_i}{T_{n,1;0.05}} \right].$$

9. Let $X_1, X_2, \ldots, X_n$ be independent and identically distributed Poisson random variables with mean $\sqrt{\lambda}$ where $\lambda > 0$.

(a) (8 pts.) Construct the *uniformly most powerful* (*UMP*) level α test of H_0: $\lambda = 1$ versus H_1: $\lambda > 1$. If $n = 1$ and $\alpha = .05$, what is the rejection region of the test?

The likelihood function for λ is

$$L(\lambda) = \prod_{i=1}^{n} \frac{1}{x_i!} (\sqrt{\lambda})^{x_i} \exp(-\sqrt{\lambda}) = \left(\prod_{i=1}^{n} \frac{1}{x_i!}\right) (\sqrt{\lambda})^{\sum_{i=1}^{n} x_i} \exp(-n\sqrt{\lambda}).$$

Let $T(x) = \sum_{i=1}^{n} x_i$. for any $\lambda_1 > \lambda_2$,

$$\frac{L(\lambda_1)}{L(\lambda_2)} = \left(\frac{\sqrt{\lambda_1}}{\sqrt{\lambda_2}}\right)^{\sum_{i=1}^{n} x_i} \exp(-n(\sqrt{\lambda_1} - \sqrt{\lambda_2}))$$

is an increasing function of $T(x)$. So $T(x)$ has a monotone likelihood ratio. By Karlin-Rubin's theorem, a UMP test has a rejection region

$$R = \{x : T(x) > k\}.$$

Under H_0, $T(X)$ has a $\text{poisson}(n)$. Then we should take k to be the smallest integer such that $P(\text{poisson}(n) > k) \le \alpha$.

If $n=1$ and $\alpha = 0.05$, the rejection region is

$$R = \{x_1 > k_0\}$$

where k_0 is the smallest integer such that $\sum_{k=0}^{k_0} \frac{1}{k!} \ge 0.95 \exp(1)$.

9. (cont)

(b) (5 pts.) Show that the maximum likelihood estimator of λ is $\hat{\lambda}_n := \bar{X}^2$ where $\bar{X} = \frac{1}{n}\sum_1^n X_i$.

Followif from part(a), the log-likelihood function for λ is

$$\ell(\lambda) = -\sum_{i=1}^{n} \log(x_i!) + \frac{1}{2}\sum_{i=1}^{n} x_i \log\lambda - n\sqrt{\lambda}.$$

$$\frac{\partial \ell(\lambda)}{\partial \lambda} = \frac{1}{2}\sum_{i=1}^{n}\frac{X_i}{\lambda} - \frac{n}{2}\lambda^{-\frac{1}{2}} = 0 \Rightarrow \hat{\lambda}_n = \bar{X}^2.$$

$$\frac{\partial^2 \ell(\lambda)}{\partial \lambda^2} = -\frac{1}{2}\sum_{i=1}^{n}\frac{X_i}{\lambda^2} + \frac{n}{4}\lambda^{-\frac{3}{2}} = -\frac{1}{2}n\bar{X}/\lambda^2 + \frac{n}{4}\lambda^{-\frac{3}{2}}.$$

$$\left.\frac{\partial^2 \ell(\lambda)}{\partial \lambda^2}\right|_{\hat{\lambda}} = -\frac{n}{2}\frac{\bar{X}}{\bar{X}^4} + \frac{n}{4}\bar{X}^{-3} = -\frac{n}{4}\bar{X}^{-3} < 0.$$

So the MLE for λ is $\hat{\lambda}_n = \bar{X}^2$.

(c) (5 pts.) Show that $\sqrt{n}(\hat{\lambda}_n - \lambda)$ converges in distribution to a normal distribution with mean 0 and variance $4\lambda^{3/2}$.

By CLT, $\sqrt{n}(\bar{X} - \sqrt{\lambda}) \xrightarrow{d} N(0, \sqrt{\lambda})$. Let $g(\bar{X}) = \bar{X}^2$.

By Delta method,

$$\sqrt{n}(g(\bar{X}) - g(\sqrt{\lambda})) \xrightarrow{d} N(0, (g'(\sqrt{\lambda}))^2\sqrt{\lambda}).$$

where $g'(\sqrt{\lambda}) = 2\sqrt{\lambda}$. Therefore,

$$\sqrt{n}(g(\bar{X}) - g(\sqrt{\lambda})) \xrightarrow{d} N(0, 4\lambda^{\frac{3}{2}}).$$

10. Two objects are to be weighed. Their true weights are β_1 and β_2. The scale is subject to error and for Object i it gives a measurement $Y_i = \beta_i + \epsilon_i, i = 1, 2$. For any true weight β, the scale returns the measurement $Y = \beta + \epsilon$ where $E(Y) = \beta$ and $V(Y) = \sigma^2$ where the variance σ^2 does not depend on β. We model the errors in repeated weighings as uncorrelated. The two objects are weighed separately and then they are weighed together.

(a) (5 pts.) Write the equations of the linear model for this experiment with $n = 3$ observations and $p = 2$ beta parameters β_1 and β_2.

$$Y_1 = \beta_1 + \epsilon_1$$
$$Y_2 = \beta_2 + \epsilon_2$$
$$Y_3 = \beta_1 + \beta_2 + \epsilon_3$$

(b) (5 pts.) Solve the normal equations to determine the least squares estimators for β_1 and β_2 and compute the variances of the estimators.

Normal Equations: $(X'X)\hat{\beta} = X'Y$ where $X = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 1 \end{bmatrix}$

$X'X = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$ $\quad X'Y = \begin{bmatrix} Y_1 + Y_3 \\ Y_2 + Y_3 \end{bmatrix}$

NE are $\quad 2\hat{\beta}_1 + \hat{\beta}_2 = Y_1 + Y_3$

$\quad\quad \hat{\beta}_1 + 2\hat{\beta}_2 = Y_2 + Y_3$

Solution

$\hat{\beta}_1 = \frac{1}{3}(2Y_1 - Y_2 + Y_3)$

$\hat{\beta}_2 = \frac{1}{3}(-Y_1 + 2Y_2 + Y_3)$

$V(\hat{\beta}_1) = \frac{1}{9}[4 + 1 + 1]\sigma^2 = \frac{2}{3}\sigma^2$

$V(\hat{\beta}_2) = \frac{1}{9}[1 + 4 + 1]\sigma^2 = \frac{2}{3}\sigma^2$

11. (4 pts.) Consider a fitted regression line of Y on X.

(a) The slope of the regression line always lies between -1 and 1. True or False

(b) The regression line always passes through $(\bar{x}, \bar{y})$. True or False

(c) The regression line always passes through the point (0, 0). True or False

(d) The intercept of the regression line always lies between -1 and 1. True or False

12. (1 pt.) A regression line of the score of Exam 2 on Exam1 for a 300 level stat course in fall of 2003 yielded the regression line y = 2.3 + 0.84x. A similar line for a 400 level stat course in fall 2004 yielded y = 2.3 + 0.61x.

(a) The correlation between the two exam scores for the 300 level courses is larger than that for the 400 level course. True or False

(b) There is not enough information to draw any conclusions about the relative sizes of the correlations. True or False