Practice Examination # 3

Sta 230: Probability

December 13, 2012

This is a closed-book exam so do not refer to your notes, the text, or any other books (please put them on the floor). You may use a single sheet of notes or formulas and a calculator, but materials may not be shared. A formula sheet and four blank worksheets are attached to the exam.

You must show your work to get partial credit. Even correct answers will not receive full credit without justification.

Please give all numerical answers to at least two correct digits or as exact fractions reduced to lowest terms. Write your solutions as clearly as possible and make sure it's easy to find your answers (circle them if necessary), since you will not receive credit for work that I cannot understand or find. Good Luck!

If you find a question confusing please **ask me** to clarify it.

Problem 1 : The United States Senate contains two senators from each of the 50 states.

a. If a committee of eight senators is selected at random, what is the probability it will contain at least one of the two senators from a specified state ?

$$\mathbf{P}(\text{at least one of two}) = 1 - \mathbf{P}(\text{none})$$
$$\mathbf{P}(\text{none}) = \frac{48 \times 47 \times \cdots 40}{50 \times 49 \times \cdots 42}$$

b. What is the probability that a group of fifty senators selected at random will contain one senator from each state ?

You can skip this problem

$$\mathbf{P} = \sum_{k=0}^{49} \frac{k}{50-k}.$$

c. Assume there are 47 republican senators and 53 democrats. What is the probability that a committee of 20 senators randomly selected will contain 13 republicans ?

$$\mathbf{P} = \frac{\binom{47}{13}\binom{53}{7}}{\binom{100}{20}}.$$

d. Assume you do not know how many democrats or republicans there are in the Senate. You observe a committee of 14 Senators that has been drawn at random contains 8 democrats. Provide a distribution over the number of Republicans in the full Senate.

Set the number of democrats in the senate to q. Set our prior probability of q to be uniform over 1 to 100 or $\pi(q) = \frac{1}{100}$

$$\mathbf{P}(k = 8 \mid q) = \frac{\binom{q}{8}\binom{100-q}{6}}{\binom{100}{14}}.$$

$$\mathbf{P}(q \mid k=8) = \frac{\frac{\binom{q}{8}\binom{100-q}{6}}{\binom{100}{14}} \times \frac{1}{100}}{\sum_{q=1}^{100} \frac{\binom{q}{8}\binom{100-q}{6}}{\binom{100}{14}} \times \frac{1}{100}}.$$

Problem 2 : A poll from the 2000 election looked at past data and forecasted that G.W. Bush had a vote share of 49.1% with variance of 2.2%.

[1.] Convert the above numbers 49.1 and 2.2 into α and β of a Beta distribution and state the Beta distribution on the parameter θ, the probability of someone in Florida voting for Bush.

$$\mathbf{E}[\theta] = .491 = \frac{\alpha}{\alpha+\beta}$$
$$\mathbf{V}[\theta] = .022 = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}.$$

Solving the above gives us $\alpha_0 = 5.0866$ and $\beta_0 = 5.2377$.

[2.] In July a polling firm asked 509 people who they would vote for and got a response of 279 for Bush and 230 for Gore. If θ is the probability of voting for Bush then state

$$p(k = 279 \mid \theta, n = 509).$$

$$p(k = 279 \mid \theta, n = 509) = \binom{509}{279}\theta^{279}(1-\theta)^{509-279}.$$

[3.] Given $p(k = 279 \mid \theta, n = 509)$ from part (2) and the prior distribution on θ given by the Beta distribution in part (1) state

$$f(\theta \mid k = 279, n = 509).$$

$$\begin{aligned} f(\theta \mid k = 279, n = 509) &= \frac{\binom{509}{279}\theta^{279}(1-\theta)^{509-279} \times \frac{\Gamma(5.0866+5.2377)}{\Gamma(5.0866)\Gamma(5.2377)}\theta^{5.0866-1}(1-\theta)^{5.2377-1}}{K} \\ &\propto \theta^{279+5.0866-1}(1-\theta)^{509-279+5.2377-1}. \end{aligned}$$

This implies that θ is distributed as $\text{Beta}(284.0866, 235.2377)$.

[4.] State $\mathbf{E}[\theta \mid k = 729, n = 509]$.

Given the above and the expectation of the Beta distribution

$$\mathbf{E}[\theta \mid k = 729, n = 509] = \frac{\alpha}{\alpha + \beta} = .547$$

Problem 3 : The pdf of X and Y is

$$f(x,y) = c(x+y^2), \quad 0 \le x \le 1, 0 \le y \le 1$$

and 0 otherwise.

Determine

1. c

$$\begin{aligned} 1 &= c\int_0^1 \int_0^1 (x+y^2)dx\,dy \\ &= c\int_0^1 \left[\int_0^1 xdx\right] + y^2)\,dy \\ &= c\int_0^1 (\frac{1}{2}+y^2)\,dy \\ &= c(1/2+1/3). \end{aligned}$$

So $c = 6/5$.

2. $f(x)$

$$\begin{aligned} f(x) &= \frac{6}{5}\int_0^1 (x+y^2)dy \\ &= \frac{6}{5}\left[x\int_0^1 dy + \int_0^1 y^2 dy\right] \\ &= \frac{6}{5}\left[x+\frac{1}{3}\right]. \end{aligned}$$

3. $f(y)$

$$\begin{aligned} f(y) &= \frac{6}{5}\int_0^1 (x+y^2)dx \\ &= \frac{6}{5}\left[y^2\int_0^1 dx + \int_0^1 xdx\right] \\ &= \frac{6}{5}\left[y^2+\frac{1}{2}\right]. \end{aligned}$$

4. $f(x \mid y)$

$$\begin{aligned} f(x \mid y) &= \frac{f(x,y)}{f(y)} \\ &= \frac{x+y^2}{\frac{1}{2}+y^2}. \end{aligned}$$

5. $f(X < 1/2 \mid Y = 1/2)$

$$\begin{aligned} f(X < 1/2 \mid Y = 1/2) &= \frac{f(x < 1/2, y = 1/2)}{f(y = 1/2)} \\ &= \frac{\frac{6}{5}\int_0^{1/2}(x+1/4)dx}{9/10}. \end{aligned}$$

6. $\mathbf{P}(X + Y < .2)$

$$\begin{aligned} \mathbf{P}(X + Y < .2) &= \frac{6}{5}\int_0^{.2}\int_0^{.2-x}(x+y^2)dxdy \\ &= \frac{6}{5}\int_0^{.2}\left[\int_0^{.2-x} xdy + \int_0^{.2-x} y^2)dy\right] dx \end{aligned}$$

7. If X and Y are independent. They are not

$$f(x,y) \neq f(x) \times f(y).$$

Problem 4: Consider the following model

$$Y_i = \beta X_i + \varepsilon_i, \quad \varepsilon_i \overset{iid}{\sim} \mathrm{N}(0,1),$$

Determine

1. The pdf $f(y_1, ..., y_n \mid x_1, ..., x_n, \beta)$.

$$f(y_1, ..., y_n \mid x_1, ..., x_n, \beta) = \left(\frac{1}{\sqrt{2\pi}}\right)^n \prod_{i=1}^{n} \left[\exp(-\frac{1}{2}\left(y_i - \beta x_i)^2\right)\right]$$

2. Assume β is unknown, maximize the above pdf to provide an estimate for β.

$$\max_{\beta} \left[f(y_1, ..., y_n \mid x_1, ..., x_n, \beta)\right] \quad = \quad \max_{\beta} \left[-\frac{1}{2}\sum_{i=1}^{n}(y_i - \beta x_i)^2\right]$$

using calculus we get $\hat{\beta} = \frac{\sum_i (x_i y_i)}{\sum_i x_i^2}$.

Now consider the following model

$$Y_i = \exp(\beta X_i + \varepsilon_i), \quad \varepsilon_i \overset{iid}{\sim} \mathrm{N}(0,1),$$

Determine

3. The pdf $f(y_1, ..., y_n \mid x_1, ..., x_n, \beta)$. Consider $z_i = \log(y_i)$. We will substitute z_i for $\log(y_i)$ but we need to consider that

$$|f(y)dy| = |f(z)dz|, \quad \frac{dz}{dy} f(z) = f(y), \quad \text{so } \frac{dz_i}{dy_i} = 1/Y_i$$

$$f(y_1, ..., y_n \mid x_1, ..., x_n, \beta) = \left(\frac{1}{\sqrt{2\pi}}\right)^n \prod_{i=1}^{n} \left[\frac{dz_i}{dy_i}\exp(-\frac{1}{2}\left(z_i - \beta x_i)^2\right)\right]$$

Problem 5 : . There are two dogs in a house. Each day the random variable $X \sim \text{Bin}(p)$ is the probability that the first dog is let outside. If the dogs are let outside then the number of barks are

$$(Y \mid X = 1) \sim \text{Pois}(\lambda_1), \quad (Y \mid X = 0) \sim \text{Pois}(\lambda_2).$$

1. If you observe over a day bark counts $Y_1, ..., Y_n$ state the probability that dog 1 was let outside that day, provide the same computation for dog 2.

$$f(\text{dog } 1 \mid Y_1, ..., Y_n) = \frac{\prod_{i=1}^n \left[\frac{\lambda_1^{y_i}}{y_i!} e^{-\lambda_1}\right] p}{\prod_{i=1}^n \left[\frac{\lambda_1^{y_i}}{y_i!} e^{-\lambda_1}\right] p + \prod_{i=1}^n \left[\frac{\lambda_2^{y_i}}{y_i!} e^{-\lambda_2}\right] (1-p)}.$$

$$f(\text{dog } 2 \mid Y_1, ..., Y_n) = \frac{\prod_{i=1}^n \left[\frac{\lambda_2^{y_i}}{y_i!} e^{-\lambda_2}\right] (1-p)}{\prod_{i=1}^n \left[\frac{\lambda_1^{y_i}}{y_i!} e^{-\lambda_1}\right] p + \prod_{i=1}^n \left[\frac{\lambda_2^{y_i}}{y_i!} e^{-\lambda_2}\right] (1-p)}.$$

2. What is $\mathbf{E}[Y] = p\lambda_1 + (1-p)\lambda_2 = \mathbf{V}[Y]$.

3. State the joint, conditional, and marginal probabilities for X, Y.

   See above for

   $$f(\text{dog } 1 \mid Y_1, ..., Y_n) \quad f(\text{dog } 2 \mid Y_1, ..., Y_n)$$

   The joints are

$$\begin{aligned} f(\text{dog } 1, Y_1, ..., Y_n) &= \prod_{i=1}^n \left[\frac{\lambda_1^{y_i}}{y_i!} e^{-\lambda_1}\right] p \\ f(\text{dog } 2, Y_1, ..., Y_n) &= \prod_{i=1}^n \left[\frac{\lambda_2^{y_i}}{y_i!} e^{-\lambda_2}\right] (1-p). \end{aligned}$$

Problem 6 : . You have a choice of placing your investments in one of six funds. The return in a year in dollars for the six firms are

$$X_1 \sim \text{Pois}(\lambda_1), \quad X_2, ..., X_6 \overset{iid}{\sim} \text{Pois}(\lambda_2).$$

1. What is the probability that at least one the funds returned $\geq \$1$? Let $Z = \max(X_1, ..., X_6)$

$$\begin{aligned} \mathbf{Pr}(Z < 1) &= \mathbf{Pr}((X_1 < 1) \cap \cdots \cap (X_6 < 1)) \\ &= \mathbf{Pr}(X_1 < 1) \times \cdots \times \mathbf{Pr}(X_6 < 1) \\ &= \left[\frac{e^{-\lambda_1}\lambda_1^0}{0!}\right]\left[\frac{e^{-\lambda_2}\lambda_2^0}{0!}\right]^5. \end{aligned}$$

   So

$$\mathbf{Pr}(Z \geq 1) = 1 - \left[\frac{e^{-\lambda_1}\lambda_1^0}{0!}\right]\left[\frac{e^{-\lambda_2}\lambda_2^0}{0!}\right]^5.$$

2. What is $\mathbf{E}[\max\{X_1, ..., X_6\}]$? We'll use the tail probability formulation. Set $Z = \max\{X_1, ..., X_6\}$

$$\mathbf{E}[Z] = \sum_t \mathbf{P}(Z \geq t).$$

   Where

$$\mathbf{P}(Z \geq t) = 1 - \left[\sum_{k=0}^{t} \frac{e^{-\lambda_1}\lambda_1^k}{k!}\right]\left[\sum_{k=1}^{t} \frac{e^{-\lambda_2}\lambda_2^k}{k!}\right]^5.$$

Problem 7 : . Suppose$X$ and Y are random variables with $\mathrm{Var}(X) = 9$, $\mathrm{Var}(Y) = 4$, and $\rho(X, Y) = -1/6$.

Determine

1. $\mathrm{Var}(X + Y)$ and $\mathrm{Var}(X - 3Y + 4)$.

$$-1/6 = \rho = \frac{\mathrm{Cov}(X, Y)}{\sqrt{\mathrm{Var}(X)\mathrm{Var}(Y)}} \Longrightarrow \mathrm{Cov}(X, Y) = \rho\sqrt{\mathrm{Var}(X)\mathrm{Var}(Y)} = -1$$

$$\mathrm{Var}(X + Y) = \mathrm{Var}(X) + \mathrm{var}(Y) + 2\mathrm{Cov}(X, Y) =$$

$$\mathrm{Var}(X-3Y+4) = \mathrm{Var}(X)+\mathrm{Var}(3Y)+2\mathrm{Cov}(X, -3Y) = \mathrm{Var}(X)+9\mathrm{Var}(Y)-6\mathrm{Cov}(X, Y)$$

2. $\mathrm{Cov}(X, Y)$.

(Nearly) Blank Work-Sheet #1 (of 4)

(Nearly) Blank Work-Sheet #2 (of 4)

(Nearly) Blank Work-Sheet #3 (of 4)

(Nearly) Blank Work-Sheet #4 (of 4)

Distributions

$$
\begin{array}{llcll}
\text{Binomial} & \Pr[X = x] & = & \binom{n}{x} p^x q^{(n-x)},\ x = 0, 1, \ldots, n & \mu = np,\ \sigma^2 = npq \\
\text{Poisson} & \Pr[X = k] & = & \lambda^k\, e^{-\lambda}/k!,\ k = 0, 1, \ldots & \mu = \lambda,\ \sigma^2 = \lambda \\
\text{Poisson process} & \Pr[X = k|t] & = & (\alpha t)^k\, e^{-\alpha t}/k!,\ k = 0, 1, \ldots & \\
\text{Geometric} & \Pr[X = x] & = & pq^x,\ y = 0, 1, 2, \ldots & \mu = q/p,\ \sigma^2 = q/p^2 \\
\text{Hypergeometric} & \Pr[X = x] & = & \frac{\binom{M}{x}\binom{N-M}{n-x}}{\binom{N}{n}},\ x = 0, 1, 2, \ldots & \mu = n\frac{M}{N} \\
 & & & & \sigma^2 = \frac{N-n}{N-1} n \frac{M}{N}(1 - \frac{M}{N}) \\
\text{Exponential} & \Pr[X \le x] & = & 1 - e^{-\lambda x},\ x \in (0, \infty) & \mu = 1/\lambda,\ \sigma^2 = 1/\lambda^2 \\
 & f(x) & = & \lambda e^{-\lambda x},\ x > 0;\ 0,\ x \le 0 & \\
\text{Uniform} & \Pr[X \le x] & = & (x - a)/(b - a),\ x \in (a, b) & \mu = \frac{a+b}{2},\ \sigma^2 = \frac{(b-a)^2}{12} \\
 & f(x) & = & \frac{1}{b-a},\ a<x<b;\ 0,\ x \notin (a, b) & \\
\text{Normal} & \Pr[X \le x] & = & \Phi(\frac{x-\mu}{\sigma}),\ x \in \mathbb{R} & \mu = \mu,\ \sigma^2 = \sigma^2 \\
 & f(x) & = & \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/2\sigma^2},\ x \in \mathbb{R} &
\end{array}
$$

$$
\begin{aligned}
P_k^n &= \frac{n!}{(n-k)!} = \overbrace{(n)(n-1)\cdots(n-k+1)}^{k \text{ terms}} \\
\text{“}n \text{ choose } k\text{”: } \binom{n}{k} &= \frac{n!}{k!(n-k)!} = \frac{(n)(n-1)\cdots(n-k+1)}{(k)(k-1)\cdots(1)}
\end{aligned}
$$

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

$$\sum_{k=0}^{\infty} ar^k = a/(1-r) \qquad \sum_{k=0}^{n} ar^k = (a - ar^{n+1})/(1-r)$$

$$\Pr[X \le x] = F(x) = \int_{-\infty}^{x} f(t)\,dt \qquad f(x) = F'(x)$$

$$\mathsf{E}[g(X)] = \int g(x) f(x)\,dx \qquad \mathsf{E}[g(X)] = \sum_x g(x) P(x)$$

$$\mu = \mathsf{E}[X] = \int_{-\infty}^{\infty} x f(x)\,dx \qquad \sigma^2 = \mathsf{E}[(X-\mu)^2] = \mathsf{E}[X^2] - \mu^2$$

A **and** $B : A \cap B$ **not** $A : A^c$ A **or** $B : A \cup B$

$$\mathsf{P}[A|B] = \mathsf{P}[A \cap B]/\mathsf{P}[B] \qquad \mathsf{P}[B|A] = \mathsf{P}[A \cap B]/\mathsf{P}[A]$$

$$\mathsf{P}[A \cap B] = \mathsf{P}[A|B]\,\mathsf{P}[B] \qquad \mathsf{P}[A \cup B] = \mathsf{P}[A] + \mathsf{P}[B] - \mathsf{P}[A \cap B]$$

Independent $\Leftrightarrow \mathsf{P}[A \cap B] = \mathsf{P}[A]\,\mathsf{P}[B]$; Exclusive $\Leftrightarrow \mathsf{P}[A \cap B] = 0$

Bayes: $\mathsf{P}[A_i|B] = \frac{\mathsf{P}[B|A_i]\mathsf{P}[A_i]}{\mathsf{P}[B|A_1]\mathsf{P}[A_1] + \ldots + \mathsf{P}[B|A_k]\mathsf{P}[A_k]}$ $A_i \cap A_j = \emptyset, \quad \mathsf{P}[A_1] + \ldots + \mathsf{P}[A_k] = 1$

Sterling: $n! \approx \sqrt{2\pi n}\left(\frac{n}{e}\right)^n$

	Continuous	Discrete
Mean	$\mu = \mathsf{E}[X] = \int_{-\infty}^{\infty} x\, f(x)dx$	$= \sum xp(x)$
Variance	$\sigma^2 = \mathsf{Var}[X] = \int_{-\infty}^{\infty} (x-\mu)^2\, f(x)dx$	$= \sum (x-\mu)^2 p(x)$
	$= \mathsf{E}[X^2] - \mathsf{E}[X]^2$	$= \mathsf{E}[X^2] - \mathsf{E}[X]^2$
Expectation	$\mathsf{E}[g(X,Y)] = \iint g(x,y)\, f(x,y)dx\, dy$	$= \sum_{x,y} g(x,y)\, p(x,y)$
Covariance	$\mathsf{Cov}[X,Y] = \iint (x-\mu_x)(y-\mu_y)\, f(x,y)dx\, dy$	$= \sum_{x,y} (x-\mu_x)(y-\mu_y)\, p(x,y)$
	$= \mathsf{E}[XY] - \mathsf{E}[X]\mathsf{E}[Y]$	$= \mathsf{E}[XY] - \mathsf{E}[X]\mathsf{E}[Y]$
Correlation coefficient	$\rho[X,Y] = \mathsf{Cov}[X,Y]/\sigma_X \sigma_Y$	
Conditional	$f(x\|y) = f(x,y)/f_2(y)$	$= p(x,y)/p_2(y)$
Marginal	$f_1(x) = \int_{-\infty}^{\infty} f(x,y)dy$	$= \sum_y p(x,y)$
	$f_2(y) = \int_{-\infty}^{\infty} f(x,y)dx$	$= \sum_x p(x,y)$
Independent	$\Leftrightarrow f(x,y) = f_1(x)\, f_2(y)$	$\Leftrightarrow p(x,y) = p_1(x)\, p_2(y)$
	$\Leftrightarrow f(x\|y) = f_1(x)$	$\Leftrightarrow p(x\|y) = p_1(x)$
	$\Leftrightarrow f(y\|x) = f_2(y)$	$\Leftrightarrow p(y\|x) = p_2(y)$

Conf. Intervals	$\mu = \bar{x} \pm z_{\alpha/2} S/\sqrt{n}$	Large Sample
	$= \bar{x} \pm t_{\alpha/2} S/\sqrt{n}$	Small Sample (normal)
	$p = \hat{p} \pm z_{\alpha/2}\sqrt{\hat{p}\,\hat{q}\,/n}$	Population Proportion
Diff. of Means	$\mu_1 - \mu_2 = (\bar{x}_1 - \bar{x}_2) \pm z_{\alpha/2}\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$	Large Samples
	$= (\bar{x}_1 - \bar{x}_2) \pm t_{\alpha/2}\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$	Small Samples
	$= \overline{x_1 - x_2} \pm z_{\alpha/2} s_\Delta/\sqrt{n}$	Matched Pairs, large n
	$= \overline{x_1 - x_2} \pm t_{\alpha/2} s_\Delta/\sqrt{n}$	Matched Pairs, small n, normal

Sample Mean	$\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i$	
Sample Variance	$s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2$	Deg fdm: $\nu = n - 1$
Regression slope	$b_1 = \hat{\beta}_1 = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2}$	
Regression offset	$b_0 = \hat{\beta}_0 = \frac{\sum y_i - \hat{\beta}_1 \sum x_i}{n} = \bar{y} - \hat{\beta}_1 \bar{x}$	
Sum Square Error	$SSE = \sum (y_i - \hat{y}_i)^2$	
Estimate of variance	$\hat{\sigma^2} = \frac{SSE}{n-2}$	
TSS	$SST = \sum (y_i - \bar{y})^2$	
CD	$r^2 = 1 - \frac{SSE}{SST}$	
F-dist	$F = \frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2}$ F-dist with $m-1$,$n-1$ d.o.f.	