Stat 342 Exam 1
Fall 2014

I have neither given nor received unauthorized assistance on this exam.

KEY

Name Signed **Date**

Name Printed

There are 10 questions on the following 5 pages. Do as many of them as you can in the available time. I will score each question out of 10 points AND TOTAL YOUR BEST 7 SCORES. (That is, this is a 70 point exam.)

1. Random variables x and y have a joint probability mass function specified in the table below.

blue is sum of x and y

		$x=1$	$x=2$	$x=3$	$x=4$	
y	2	3 .08	4 .10	5 .12	6 .10	.4
	1	2 .06	3 .04	4 .11	5 .09	.3
	0	1 .06	2 .06	3 .07	4 .11	.3
		.2	.2	.3	.3	

10 pts **a)** Find the SEL optimal predictor of y based on x, say $\hat{y}^{\text{opt}}_{\text{SEL}}(x)$. (Specify the 4 values $\hat{y}^{\text{opt}}_{\text{SEL}}(1), \hat{y}^{\text{opt}}_{\text{SEL}}(2), \hat{y}^{\text{opt}}_{\text{SEL}}(3)$, and $\hat{y}^{\text{opt}}_{\text{SEL}}(4)$. There is no need to do the arithmetic.)

We want the conditional mean function. That is

for $x=1$ $\quad \hat{y}(1) = 0\left(\frac{.06}{.2}\right) + 1\left(\frac{.06}{.2}\right) + 2\left(\frac{.08}{.2}\right)$

$x=2$ $\quad \hat{y}(2) = 0\left(\frac{.06}{.2}\right) + 1\left(\frac{.04}{.2}\right) + 2\left(\frac{.1}{.2}\right)$

$x=3$ $\quad \hat{y}(3) = 0\left(\frac{.07}{.3}\right) + 1\left(\frac{.11}{.3}\right) + 2\left(\frac{.12}{.3}\right)$

$x=4$ $\quad \hat{y}(4) = 0\left(\frac{.11}{.3}\right) + 1\left(\frac{.09}{.3}\right) + 2\left(\frac{.10}{.3}\right)$

10 pts **b)** Find the 0-1 loss predictor of y based on x, say $\hat{y}^{\text{opt}}_{\text{0-1}}(x)$. (Specify the 4 values $\hat{y}^{\text{opt}}_{\text{0-1}}(1), \hat{y}^{\text{opt}}_{\text{0-1}}(2), \hat{y}^{\text{opt}}_{\text{0-1}}(3)$, and $\hat{y}^{\text{opt}}_{\text{0-1}}(4)$.)

Here we want the conditional mode function. That is

for $x=1$ $\quad \hat{y}(1) = 2$

$x=2$ $\quad \hat{y}(2) = 2$

$x=3$ $\quad \hat{y}(3) = 2$

$x=4$ $\quad \hat{y}(4) = 0$

10 pts **c)** Suppose that one begins a Gibbs SSS algorithm at $(x^0, y^0) = (1,1)$. Give the distribution one uses to generate x^1. Then, supposing that $x^1 = 2$, give the distribution that one uses in order to generate y^1.

For generating x^1 we use the conditional distribution of x given $y^1 = 1$ i.e.

x	1	2	3	4
$f(x\|1)$	$\frac{.06}{.3}$	$\frac{.04}{.3}$	$\frac{.11}{.3}$	$\frac{.09}{.3}$

For generating y^1 we use the conditional dsn of y given $x=2$ i.e.

y	0	1	2
$f(y\|2)$	$\frac{.06}{.2}$	$\frac{.04}{.2}$	$\frac{.10}{.2}$

10 pts **d)** Find the distribution of the sum $t = x + y$. (List possible values and corresponding probabilities in a table below.)

Looking at the table for sums and probabilities we see that possible values are 1, 2, 3, 4, 5, 6 and adding probabilities we get:

t	1	2	3	4	5	6
$p(t)$	.06	.12	.19	.32	.21	.10

10 pts **2.** In the planning of a large engineering project, the number of days required to complete one step in the project is modeled as a continuous random variable with cumulative distribution function

$$F(x)=\begin{cases} 0 & \text{if } x<100 \\ \dfrac{(x-100)^3}{(100)^3} & \text{if } 100\le x\le 200 \\ 1 & \text{if } x>200 \end{cases}$$

a) A simulation is going to be used to study the feasibility of the project and variables with this distribution function are needed. For $U\sim \mathrm{U}(0,1)$ give a function $h(u)$ for which $h(U)$ has the target distribution (has cdf $F(x)$). For $u\in(0,1)$

$F^{-1}(u)$ = solution to $u=\dfrac{(x-100)^3}{100^3}$

This is $100^3 u=(x-100)^3$

$\sqrt[3]{(100)^3 u} = x-100$

$x = 100+\sqrt[3]{(100)^3 u}$

So use $F^{-1}(u) = 100+100\sqrt[3]{u}$

10 pts **b)** Approximate the probability that the sample mean of $n=16$ simulated values from the distribution specified by F exceeds 180 days. (You may use without proof the fact that the distribution has mean 175 and standard deviation 19.4.)

$\mu_{\bar{x}}=175$ and $\sigma_{\bar{x}}=\dfrac{\sigma}{\sqrt{n}}=\dfrac{19.4}{\sqrt{16}}=4.85$ and the CLT says that the standardized version of $\bar{x}$ is approximately standard normal. So

$$P[\bar{x}>180]=P\left[\frac{\bar{x}-175}{4.85}>\frac{180-175}{4.85}\right]$$

$$\approx 1-\Phi\left(\frac{180-175}{4.85}\right)$$

10 pts **3.** The exponential distribution with mean μ has standard deviation μ. If $x_1, x_2, \ldots$ are iid according to this distribution, what is an approximate distribution for

$$\ln\left(\bar{x}_n\right)$$

(the natural logarithm of the sample mean of the first n of these variables)?

$\bar{x}_n$ is approximately $N\left(\mu, \frac{\mu^2}{n}\right)$. Then $\ln(\bar{x}_n)$ is approximately normal with mean $\ln(\mu)$ and std deviation

$$\left|\frac{d}{dy}\ln(y)\Big|_{\mu}\right|\left(\frac{\mu}{\sqrt{n}}\right)$$

$$= \left|\frac{1}{\mu}\right|\left(\frac{\mu}{\sqrt{n}}\right)$$

$$= \frac{1}{\sqrt{n}}$$ (Notice BTW that this is free of μ)

10 pts **4.** If $x_1, x_2, \ldots, x_n$ are iid $\mathrm{N}\left(\mu, \sigma^2\right)$ then we know that $\frac{(n-1)s^2}{\sigma^2} \sim \chi^2_{n-1}$. Let # stand for the lower 2.5% point of the χ^2_{n-1} distribution and ## stand for the upper 2.5% point of the χ^2_{n-1} distribution. Use these elements and produce 95% confidence limits for σ. That is, give formulas for random variables L and U so that in this context

$$P\left[L < \sigma < U\right] = .95$$

for all possible $\left(\mu, \sigma^2\right)$ pairs. (Explain/make clear why your limits will work.)

$$P\Big[\underbrace{\# < \frac{(n-1)s^2}{\sigma^2} < \#\#}\Big]$$

$$\underbrace{\frac{(n-1)s^2}{\#\#} < \sigma^2 < \frac{(n-1)s^2}{\#}}$$

$$\sqrt{} < \sqrt{} < \sqrt{}$$

So 95% confidence limits for σ are

$$L = s\sqrt{\frac{n-1}{\#\#}} \quad \text{and} \quad U = s\sqrt{\frac{n-1}{\#}}$$

10 pts

5. The Poisson(λ) pmf is $f(x|\lambda)=\lambda^x \exp(-\lambda)/x!$ for non-negative integer x. The $\Gamma(\alpha,\beta)$ pdf is $g(\theta)=\theta^{\alpha-1}\exp(-\theta/\beta)/\beta^{\alpha}\Gamma(\alpha)$ for $\theta>0$. (You may use this information in the following.)

In a Bayes model, $x_1,x_2,\ldots,x_n$ are iid Poisson(λ) and *a priori* $\lambda \sim \text{Exp}(1)$. Identify the posterior distribution.

The likelihood times prior is

$$\prod_{i=1}^{n} \frac{\lambda^{x_i}\exp(-\lambda)}{x_i!} \cdot \exp(-\lambda)$$

as a function of λ this is proportional to

$$\lambda^{\sum x_i} \exp(-(n+1)\lambda)$$

This is proportional to the Γ density for λ with parameters $\alpha=\sum x_i+1$ and $\beta=\frac{1}{n+1}$

10 pts

6. A single random observation x has a distribution with pdf $f(x|\theta)=\left(x^{\theta}(\theta+1)\right)I[0<x<1]$. Find the maximum likelihood estimator of θ in this statistical model and write out (but do NOT attempt to evaluate/simplify) a completely specified definite integral that gives the squared error loss risk function of this estimator, $R(\theta)$.

$f(x|\theta)=(\theta+1)x^{\theta}$ and $\ln f(x|\theta)=\theta\ln x+\ln(\theta+1)$

So $\frac{d}{d\theta}\ln f(x|\theta)=\ln x+\frac{1}{\theta+1}$ and at a maximum of $f(x|\theta)$ or $\ln f(x|\theta)$ over θ, $\frac{d}{d\theta}\ln f(x|\theta)=0$

i.e. $\theta+1=\frac{-1}{\ln x}$ i.e. $\theta=-\frac{1}{\ln x}-1$. So $\hat{\theta}^{MLE}=-\frac{1}{\ln x}-1$

This estimator has risk function

$$R(\theta)=\int_0^1\left(-\frac{1}{\ln x}-1-\theta\right)^2(\theta+1)x^{\theta}\,dx$$