Program in Applied Mathematics
PROBABILITY AND STATISTICS PRELIMINARY EXAMINATION
January 2017

Notice: Do four of the following five problems. Place an X on the line opposite the number of the problem that you are **NOT** submitting for grading. Please do not write your name anywhere on this exam. You will be identified only by your student number, given below and on each page submitted for grading. Show all relevant work.

1. ____
2. ____
3. ____
4. ____
5. ____

Total ____

Student Number ____________________________

1. If $Z_1, Z_2, Z_3, \ldots$ are i.i.d. $Normal(0,1)$ then the random variable $S_d := \sum_{i=1}^{d} Z_i^2$, with $d \geq 1$, is said to have a chi-squared distribution with d-degrees of freedom; in short, we write $S_d \sim \chi^2(d)$. Let $f_d(t)$ denote the p.d.f. of S_d.

   (a) Deduce f_1.

   (b) For $0 < k < d$, explain why $f_d = (f_k * f_{d-k})$ i.e. $f_d(t) = \int_{-\infty}^{+\infty} f_k(s) \cdot f_{d-k}(t-s)\, ds$.

   (c) Show that $f_{2n}(t) = \frac{t^{n-1}\, e^{-t/2}}{(n-1)!\, 2^n}$, for all $t \geq 0$ and integer $n > 0$.

   (d) Recall that $\Gamma(x) = \int_0^\infty t^{x-1} e^{-t}\, dt$ for $x > 0$. Show that $f_{2n+1}(t) = \frac{t^{n-1/2}\, e^{-t/2}}{\Gamma(n+1/2)\, 2^{n+1/2}}$, for all $t \geq 0$ and integer $n > 0$.

2. In this problem, all random variables are assumed to be $\mathbb{N}$-valued i.e. take non-negative integer values. The *total variation distance* between two random variables X and Y is

$$d(X,Y) := \frac{1}{2} \sum_{k=0}^{\infty} |P(X=k) - P(Y=k)|.$$

   (In effect, this is really a distance between their distributions.)

   Consider the set $B := \{k \in \mathbb{N} \text{ such that } \mathrm{P}(X=k) \geq \mathrm{P}(Y=k)\}$. Show that:

   (a) $d(X,Y) = \mathrm{P}(X \in B) - \mathrm{P}(Y \in B)$.

   (b) $|\mathrm{P}(X \in A) - \mathrm{P}(Y \in A)| \leq \mathrm{P}(X \in AB) - \mathrm{P}(Y \in AB) + \mathrm{P}(Y \in AB^c) - \mathrm{P}(X \in AB^c)$, for all $A \subset \mathbb{N}$, where B^c is the complement of the set B and $AB^c = (A \cap B^c)$.

   (c) Use (a)-(b) to conclude that: $d(X,Y) = \max_{A \subset \mathbb{N}} |\mathrm{P}(X \in A) - \mathrm{P}(Y \in A)|$.

   Recall that $Bernoulli(p) = Binomial(1,p)$. It can be shown (take these two facts for granted) that $d\big(Bernoulli(p), Poisson(p)\big) = p(1-e^{-p})$ and that, if $X_1, \ldots, X_n$ are independent and $Y_1 \ldots, Y_n$ are independent then $d\big(\sum_{i=1}^n X_i, \sum_{i=1}^n Y_i\big) \leq \sum_{i=1}^n d(X_i, Y_i)$. The final goal of this problem is to show the celebrated *Poisson approximation of the Binomial*:

   Theorem. *If $X \sim Binomial(n,p)$ and $Y \sim Poisson(np)$ then, for each $A \subset \mathbb{N}$, $|\mathrm{P}(X \in A) - \mathrm{P}(Y \in A)| \leq np^2$.*

   (d) Show the theorem!

3. Let X_1, X_2, ..., X_n be i.i.d. random variables with normal distribution $N(\mu, \sigma^2)$. Suppose σ is known, and we want to test the hypothesis

$$H_0 : \mu = \mu_0 \quad \text{against} \quad H_1 : \mu \neq \mu_0.$$

   (a) Find the maximum likelihood estimator (MLE) for μ.

   (b) Find the generalized likelihood ratio (GLR) for the test. Call it Λ.

   (c) Suppose we reject H_0 when $-2 \ln \Lambda$ is larger than a given constant $C > 0$. Write down the critical region of the test in terms of the MLE in (a).

   (d) What kind of distribution (exact, not asymptotic) does $-2 \ln \Lambda$ follow given that H_0 is true?

4. Let X_1, X_2, ..., X_n be i.i.d. random variables with the probability density function

$$f(x; \theta) = \frac{1}{\theta^2} x e^{-\frac{x}{\theta}} I_{(0,\infty)}(x),$$

where $\theta > 0$ and $I_{(0,\infty)}(x)$ is the indicator function that takes the value 1 when $x \in (0, \infty)$ and 0 otherwise.

   (a) Find a one-dimensional sufficient statistic for the model, and find an unbiased estimator for θ^n based on this sufficient statistic.

   (b) Find the Cramér-Rao lower bound for the variance of all unbiased estimators of θ^n.

   (c) Find the variance of the estimator you found in (a). Does it attain the Cramér-Rao lower bound?

5. Let $0 < \lambda, \mu < +\infty$. Imagine a bank with a single teller. Assume customers arrive according to a homogenous Poisson point process with rate λ, and the teller requires an exponential amount of time with mean $1/\mu$ to serve each customer. Upon being served, however, customers are forced back at the end of the line with probability $0 \leq b < 1$, and otherwise leave the bank entirely.

   Let X_t denote the number of people in the bank (i.e. in line or being served) at time t. Clearly, $X = (X_t)_{t \geq 0}$ is a time-homogeneous Markov process. Based on this, respond:

   (a) What condition must (λ, μ, b) satisfy in order for X to have a stationary distribution?

   (b) Determine the stationary distribution π when the above condition is satisfied.

   (c) In equilibrium, what fraction of the time is the teller busy with customers?