May 2000

2000 Honors Examination in Statistics

Department of Mathematics and Statistics
Swarthmore College

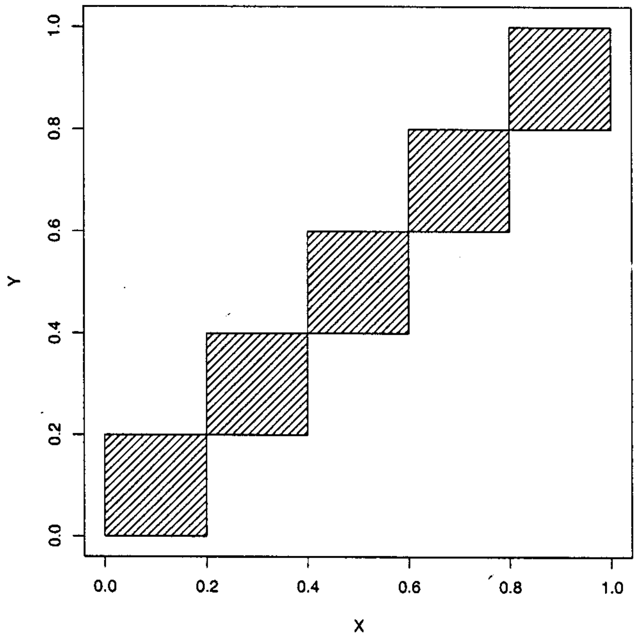
Instructions: This exam consists of a total of four questions. Number the questions clearly in your work and start each question on a new page. You must show your work to make it clear how you obtained your answers. Answers without any work may lose credit even if they are correct, and will receive no credit if incorrect.

This is a closed-book three-hour exam. You may not refer to your notes or textbooks.

1. For fixed positive integer K, suppose X and Y have the joint probability density function

$$f(x,y) = \begin{cases} c & \text{if both } \frac{k}{K} < x \le \frac{k+1}{K} \text{ and} \\ & \frac{k}{K} < y \le \frac{k+1}{K},\ k = 0, 1, \ldots, K-1 \\ 0 & \text{otherwise} \end{cases}$$

For example, below is the region over which the joint density is non-zero when $K = 5$.



(a) Find the value of c (as a function of K) that makes $f(x,y)$ a proper density function.

(b) Determine the marginal distribution of X.

(c) Find the correlation between X and Y as a function of K. What happens to the correlation as K becomes large? You may use the fact that

$$\sum_{i=1}^{n} i = \frac{n(n+1)}{2}, \qquad \sum_{i=1}^{n} i^2 = \frac{n(n+1)(2n+1)}{6}$$

2. In preparation for watching a movie at home, you are about to make some fresh popcorn. The instructions on the popcorn package say: "Under a medium flame, the time in minutes for each kernel to pop follows an exponential distribution with parameter $\ln \frac{3}{2}$ (corresponding to an average time of approximately 2.47 minutes). Times for kernels to pop are independent." Suppose you plan to pop n kernels.

For the following, write down expressions in simplest form as functions of n.

(a) What is the probability that at least one kernel pops within one minute?

(b) What is the variance of the number of kernels taking between 1 and 2 minutes to pop?

(c) Suppose that all n kernels popped within 5 minutes. Given this information, what is the probability that exactly k kernels popped within 3 minutes, where $k < n$?

(d) Assuming n is large, what is the approximate distribution of the sample mean time that n kernels pop? Be specific.

3. Let X be a random variable with cumulative distribution function (cdf) F_X, and suppose $X_1, \ldots, X_n$ are a random sample of size n.

(a) How might you estimate $\mathrm{E}(e^{-2X})$ from the sample if you did not know the specific form of F_X? Determine an expression for the estimate.

(b) Suppose $Y = \min(X_1, \ldots, X_n)$. Find a closed-form expression for the cdf of Y in terms of F_X.

(c) Determine the maximum likelihood estimate of $\mathrm{E}(e^{-2X})$ if X has the cdf $F_X(x) = 1 - e^{-\lambda x}$, for $x \geq 0$ and $\lambda > 0$. *Hint: First determine* $E(e^{-2X})$ *as a function of* λ.

(d) Suppose again that $F_X(x) = 1 - e^{-\lambda x}$, for $x \geq 0$ and $\lambda > 0$, and you are interested in testing

$\mathbf{H}_o$: $\mathrm{E}(e^{-2X}) = 1/3$
$\mathbf{H}_a$: $\mathrm{E}(e^{-2X}) = 1/2$

Derive the procedure for carrying out a most powerful test. That is, determine the test statistic (in its simplest form), state or derive its distribution, and briefly explain how to compute a p-value for the test.

4. Suppose a random variable, X, has a Poisson distribution with mean μ. Suppose $X_1, \ldots, X_n$ is a random sample of size n.

(a) Find the Bayes estimator of μ from the sample under squared-error loss, that is, the posterior mean of μ, with prior distribution

$$f(\mu) = e^{-\mu} \quad , \quad \mu > 0.$$

(b) Consider the estimator

$$T = \frac{1}{n} \sum_{i=1}^{n} X_i^2$$

Show that this estimator is never unbiased for μ.

(c) Suppose you are interested in testing the hypotheses

$\mathbf{H}_o$: $\mu = 10$
$\mathbf{H}_a$: $\mu = 9$

Derive an expression for the necessary sample size, n, to guarantee a significance level of $\alpha = 0.10$ and power of 0.90, assuming that n is large enough to permit a normal approximation to $\sum_{i=1}^{n} X_i$. You do not need to perform the computation, though your answer should be an expression that only involves numbers. You may use the fact that $\mathrm{P}(Z < 1.645) = 0.90$ for $Z \sim \mathrm{N}(0, 1)$.

Some Common Probability Distributions

Discrete probability distributions

Name	Abbreviation	Probability mass function, $p(x)$	Mean	Variance	Moment Generating Fnc, $M(t)$
Bernoulli	$\mathrm{Ber}(p)$	$p(x) = p^x(1-p)^{1-x},\ x = 0, 1$	p	$p(1-p)$	$M(t) = pe^t + 1 - p$
Binomial	$\mathrm{B}(n, p)$	$p(x) = \binom{n}{x}p^x(1-p)^{n-x},\ x = 0, 1, \ldots, n$	np	$np(1-p)$	$M(t) = (pe^t + 1 - p)^n$
Poisson	$\mathrm{Po}(\lambda)$	$p(x) = e^{-\lambda}\lambda^x/x!,\ x = 0, 1, \ldots$	λ	λ	$M(t) = \exp(\lambda(e^t - 1))$
Geometric	Ge(p)	$p(x) = p(1-p)^{x-1},\ x = 1, 2, \ldots$	$1/p$	$(1-p)/p^2$	$M(t) = \frac{pe^t}{1-(1-p)e^t}$
Negative Binomial	$\mathrm{NB}(r, p)$	$p(x) = \binom{x-1}{r-1}p^r(1-p)^{x-r},\ x = r, r+1, \ldots$	r/p	$r(1-p)/p^2$	$M(t) = \left[\frac{pe^t}{1-(1-p)e^t}\right]^r$
Hypergeometric	$\mathrm{HG}(N, M, n)$	$p(x) = \binom{M}{x}\binom{N-M}{n-x}/\binom{N}{n}$	nM/N	$\frac{nM}{N}\left[\frac{(n-1)(M-1)}{N-1} + (1 - \frac{nM}{N})\right]$	No closed form expression

Continuous probability distributions

Name	Abbreviation	Probability density function, $f(x)$	Mean	Variance	Moment Generating Fnc, $M(t)$
Uniform	$\mathrm{U}(a, b)$	$f(x) = 1/(b-a),\ a < x < b$	$(a+b)/2$	$(b-a)^2/12$	$M(t) = (e^{tb} - e^{ta})/(t(b-a))$
Exponential	$\mathrm{Ex}(\lambda)$	$f(x) = \lambda e^{-\lambda x},\ x \geq 0$	$1/\lambda$	$1/\lambda^2$	$M(t) = \lambda/(\lambda - t)$
Gamma	$\mathrm{Gam}(s, \lambda)$	$f(x) = \lambda^s e^{-\lambda x} x^{s-1}/\Gamma(s),\ x \geq 0$	s/λ	s/λ^2	$M(t) = (\lambda/(\lambda - t))^s$
Normal	$\mathrm{N}(\mu, \sigma^2)$	$f(x) = \frac{1}{\sqrt{2\pi}\sigma}\exp(\frac{-(x-\mu)^2}{2\sigma^2}),\ -\infty < x < \infty$	μ	σ^2	$M(t) = \exp(\mu t + \sigma^2 t^2/2)$
Cauchy	$\mathrm{Cau}(\theta)$	$f(x) = \frac{1}{\pi(1+(x-\theta)^2)},\ -\infty < x < \infty$	undefined	undefined	undefined