IE 230 Seat # ________ Name ___ < KEY > ___

Closed book and notes. 60 minutes.

Cover page and four pages of exam.
Pages 8 and 12 of the Concise Notes.
No calculator. No need to simplify answers.

This test is cumulative, with emphasis on Section 4.7
through Chapter 6 of Montgomery and Runger, fourth edition.

Remember: A statement is true only if it is always true.

One point: On the cover page, circle your family name.

One point: On every page, write your name.

The random vector $(X_1, X_2, \ldots, X_k)$ has a *multinomial distribution* with joint pmf

$$P(X_1 = x_1, X_2 = x_2, \ldots, X_k = x_k) = \frac{n!}{x_1!\, x_2! \cdots x_k!}\, p_1^{x_1}\, p_2^{x_2} \cdots p_k^{x_k}$$

when each x_i is a nonnegative integer and $x_1 + x_2 + \cdots + x_k = n$; zero elsewhere.

The linear combination $Y = c_0 + c_1 X_1 + c_2 X_2 + \cdots + c_n X_n$ has mean and variance

$$E(Y) = c_0 + \sum_{i=1}^{n} E(c_i X_i) = c_0 + \sum_{i=1}^{n} c_i\, E(X_i)$$

and

$$V(Y) = \sum_{i=1}^{n}\sum_{j=1}^{n} \mathrm{cov}(c_i X_i, c_j X_j) = \sum_{i=1}^{n}\sum_{j=1}^{n} c_i\, c_j\, \mathrm{cov}(X_i, X_j)$$

$$= \sum_{i=1}^{n} c_i^2\, V(X_i) + 2\sum_{i=1}^{n-1}\sum_{j=i+1}^{n} c_i\, c_j\, \mathrm{cov}(X_i, X_j)\,.$$

$\mathrm{Cov}(X, Y) = E[(X - \mu_X)(Y - \mu_Y)]$

$\mathrm{Corr}(X, Y) = \mathrm{Cov}(X, Y) / (\sigma_X\, \sigma_Y)$

$\bar{X} = \sum_{i=1}^{n} X_i / n$

$S^2 = \sum_{i=1}^{n} (X_i - \bar{X})^2 / (n - 1)$

Order statistics satisfy $X_{(1)} \le X_{(2)} \le \cdots \le X_{(n)}$.

Score ____________________________

Closed book and notes. 60 minutes.

1. (3 points each) True or false.

Consider two continuous random variables X and Y with probability density functions f_X and f_Y, expected values μ_X and μ_Y, standard deviations σ_X and σ_Y, and correlation $\rho_{X,Y}$.

(a) T F ← $\mu_X(v) = \int_{-\infty}^{\infty} f_X(v)\, dv$

(b) T ← F $f_{X,Y}(x,y) = f_Y(y)\, f_{X \mid Y=y}(x)$

(c) T ← F $\mathrm{E}(X - Y) = \mu_X - \mu_Y$

(d) T F ← $\mathrm{Var}(X - Y) = \sigma_X^2 - \sigma_Y^2$

(e) T ← F $|\mathrm{Cov}(X, Y)| \le \sigma_X\, \sigma_Y$

(f) T F ← If $\rho_{X,Y} = 0$, then X and Y are independent.

(g) T ← F If X and Y are independent, then $\rho_{X,Y} = 0$.

(h) T ← F If (X,Y) is bivariate normal, then both X and Y are normal.

2. (3 points each) Consider the notation from Question 1 above. For each of the following, indicate whether the expression is a constant, an event, a random variable, or undefined. (A constant has the same numerical value for every replication of the experiment.)

(a) $\sigma_X\, \sigma_Y$ constant ← event random variable undefined

(b) $\mu_X\, X$ constant event random variable ← undefined

(c) $X^{1/2} / Y^{1/2}$ constant event random variable ← undefined

(d) $X > 1/2$ constant event ← random variable undefined

(e) $XY > \sigma_X\, \sigma_Y$ constant event ← 0random variable undefined

3. (Problem 4–75, Montgomery and Runger, fourth edition) Assume that hits to a web site follow a Poisson process with rate 10,000 per day.

   (a) (6 points) Write the cumulative distribution function of the time between adjacent hits. (Be specific.)

   The time between hits is exponential with rate $\lambda = 10{,}000$ hits per day.
   From Page 12 of the Concise Notes, $F(t) = 1 - \exp(-\lambda t)$ if $0 \le t$ and zero otherwise.

   (b) (8 points) Consider the experiment of choosing a random day. Sketch a normal bell curve to approximate the (Poisson) distribution of the number of hits. Label and scale the horizontal axis. On your sketch, indicate the probability of more than 20,000 hits.

   From Page 8 of the Concise Notes, the Poisson mean and variance and variance are equal. Therefore, both have the numerical value 10,000. Therefore the normal approximation is based on a mean of 10,000 and standard deviation of 100 hits.

   Sketch the usual bell curve. Center at $\mu = 10{,}000$. Points of inflection at 9,900 and 10,100. Tails disappearing at about 9,700 and 10,300.

   We are asked for the probability of more than 20,000 hits. In a continuous approximation, the continuity correction suggests thinking of more than 20,000.5 hits, but that error is negligible in this example. Typically, we would shade the area to the right of 20,000 or 20,000.5, the area to the right of 100 standard deviations above the mean of 10,000. In this case, the area is not visible; the probability is incredibly close to zero.

   ---

4. (Problem 5–60, Montgomery and Runger, fourth edition) Consider the experiment of selecting one door and one door casing. The width of a door casing, X, has a mean of 24 inches and a standard deviation of 1/8 inch. The width of a door, Y, has a mean of 23.875 inches and standard deviation 1/16 inch. Their correlation is $\rho_{X,Y} = 0.7$. We are interested in the gap between the door and the casing, $G = X - Y$.

   (a) (3 points)　T　F ←　If X and Y are exponentially distributed and independent, then G has an exponential distribution.

   (b) (3 points)　T　F ←　If the assembly process could be changed so that X and Y were independent, then the variance of G would be smaller.

   (c) (3 points)　T ←　F　If the manufacturing process could be changed so that the variance of Y could be reduced, then the variance of G would be smaller.

5. Consider a data set having order statistics $x_{(1)}, x_{(2)}, \ldots, x_{(n)}$.

(a) (3 points) T ← F The sample mean is $\bar{x} = \dfrac{x_{(1)} + x_{(2)} + \cdots + x_{(n)}}{n}$.

(b) (3 points) T F ← If $n = 11$, then the median is x_6.

(c) (4 points) A stem-and-leaf plot is most like a (choose one)

(i) histogram. ← (ii) cdf. (iii) joint distribution. (iv) conditional distribution.

(d) (8 points) Suppose that $n = 9$. Sketch the empirical cumulative distribution function (cdf) by plotting $x_{(i)}$ versus $i/(n+1)$. Label and scale the axes.

Sketch two perpendicular axes.
Label the horizontal axis with x and the vertical with $\hat{F}_X(x)$.
On the horizontal axis, indicate any nine points.
From left to right, label them $x_{(1)}$ through $x_{(9)}$.
Scale the vertical axis with zero and one.
Indicate the locations of $1/10, 2/10, \ldots, 9/10$.
Fill in the points $(x_{(i)}, i/10)$.

---

7. (Problem 5–48, Montgomery and Runger, fourth edition) In an acid-base titration, a base or acid is gradually added to the other until they have completely neutralized each other. Let X and Y denote the number of milliliters of acid and of base needed for equivalence. Assume that (X, Y) is bivariate normal with $\mu_X = 120$, $\mu_Y = 100$, $\sigma_X = 5$, $\sigma_Y = 2$, and $\rho_{X,Y} = 0.85$.

(a) (3 points) Describe the experiment.

Perform one acid-base titration.

(b) (3 points) T ← F $P(X > 120 \mid Y > 104) > 0.5$.

(c) (4 points) Write the pmf or pdf of X. (Be specific.)

X is normal with mean $\mu_X = 120$ and standard deviation $\sigma_X = 5$.
From Page 12 of the Concise Notes, for every real number x,

$$f_X(x) = \frac{\exp[-(1/2)((x-120)/5)^2]}{(2\pi)^{1/2}\ 5}$$

(d) (8 points) Sketch a scatter plot showing what twenty or thirty observations of (X, Y) might look like. Label and scale all axes.

Sketch two perpendicular axes.
Label the horizontal axis with x and the vertical with y.
"Randomly" make twenty or thirty marks, roughly in an elliptical cloud.
The cloud should be narrow, with a positive slope.
On the axes, mark the means, $\mu_X = 120$ and $\mu_Y = 100$,
near the center of the cloud.
On the axes, indicate the standard deviations so that the cloud ends
about two standard deviations above the means.

Discrete Distributions: Summary Table

random variable	distribution name	range	probability mass function	expected value	variance
X	general	$x_1, x_2, \ldots, x_n$	$P(X = x)$ $= f(x)$ $= f_X(x)$	$\sum_{i=1}^{n} x_i f(x_i)$ $= \mu = \mu_X$ $= E(X)$	$\sum_{i=1}^{n} (x_i - \mu)^2 f(x_i)$ $= \sigma^2 = \sigma_X^2$ $= V(X)$ $= E(X^2) - \mu^2$
X	discrete uniform	$x_1, x_2, \ldots, x_n$	$1/n$	$\sum_{i=1}^{n} x_i / n$	$[\sum_{i=1}^{n} x_i^2 / n] - \mu^2$
X	equal-space uniform	$x = a, a+c, \ldots, b$ where	$1/n$ $n = (b-a+c)/c$	$\frac{a+b}{2}$	$\frac{c^2(n^2-1)}{12}$
"# successes in 1 Bernoulli trial"	indicator variable	$x = 0, 1$	$p^x(1-p)^{1-x}$	p where	$p(1-p)$ $p = P(\text{"success"})$
"# successes in n Bernoulli trials"	binomial	$x = 0, 1, \ldots, n$	$C_x^n p^x (1-p)^{n-x}$	np where	$np(1-p)$ $p = P(\text{"success"})$
"# successes in a sample of size n from a population of size N containing K successes"	hyper-geometric (sampling without replacement)	$x =$ $(n-(N-K))^+$, ..., $\min\{K, n\}$ and integer	$C_x^K C_{n-x}^{N-K} / C_n^N$	np where	$np(1-p)\frac{(N-n)}{(N-1)}$ $p = K/N$
"# Bernoulli trials until 1st success"	geometric	$x = 1, 2, \ldots$	$p(1-p)^{x-1}$	$1/p$ where	$(1-p)/p^2$ $p = P(\text{"success"})$
"# Bernoulli trials until rth success"	negative binomial	$x = r, r+1, \ldots$	$C_{r-1}^{x-1} p^r (1-p)^{x-r}$	r/p where	$r(1-p)/p^2$ $p = P(\text{"success"})$
"# of counts in time t from a Poisson process with rate λ"	Poisson	$x = 0, 1, \ldots$	$e^{-\mu} \mu^x / x!$	μ where	μ $\mu = \lambda t$

Result. For $x = 1, 2, \ldots$, the geometric cdf is $F_X(x) = 1 - (1-p)^x$.

Result. The geometric distribution is the only discrete memoryless distribution. That is, $P(X > x + c \mid X > x) = P(X > c)$.

Result. The binomial distribution with $p = K/N$ is a good approximation to the hypergeometric distribution when n is small compared to N.

Continuous Distributions: Summary Table

random variable	distribution name	range	cumulative distrib. func.	probability density func.	expected value	variance
X	general	$(-\infty, \infty)$	$P(X \le x)$ $= F(x)$ $= F_X(x)$	$\left.\frac{dF(y)}{dy}\right\|_{y=x}$ $= f(x)$ $= f_X(x)$	$\int_{-\infty}^{\infty} x f(x) dx$ $= \mu = \mu_X$ $= E(X)$	$\int_{-\infty}^{\infty} (x-\mu)^2 f(x) dx$ $= \sigma^2 = \sigma_X^2$ $= V(X)$ $= E(X^2) - \mu^2$
X	continuous uniform	$[a, b]$	$\frac{x-a}{b-a}$	$\frac{1}{b-a}$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$
X	triangular	$[a, b]$	$(x-a)f(x)/2$ if $x \le m$, else $1-(b-x)f(x)/2$	$\frac{2(x-d)}{(b-a)(m-d)}$ ($d = a$ if $x \le m$, else $d = b$)	$\frac{a+m+b}{3}$	$\frac{(b-a)^2-(m-a)(b-m)}{18}$
sum of random variables	normal (or Gaussian)	$(-\infty, \infty)$	Table III	$\frac{e^{\frac{-1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}}{\sqrt{2\pi}\,\sigma}$	μ	σ^2
time to Poisson count 1	exponential	$[0, \infty)$	$1 - e^{-\lambda x}$	$\lambda e^{-\lambda x}$	$1/\lambda$	$1/\lambda^2$
time to Poisson count r	Erlang	$[0, \infty)$	$\sum_{k=r}^{\infty} \frac{e^{-\lambda x}(\lambda x)^k}{k!}$	$\frac{\lambda^r x^{r-1} e^{-\lambda x}}{(r-1)!}$	r/λ	r/λ^2
lifetime	gamma	$[0, \infty)$	numerical	$\frac{\lambda^r x^{r-1} e^{-\lambda x}}{\Gamma(r)}$	r/λ	r/λ^2
lifetime	Weibull	$[0, \infty)$	$1 - e^{-(x/\delta)^\beta}$	$\frac{\beta x^{\beta-1} e^{-(x/\delta)^\beta}}{\delta^\beta}$	$\delta\Gamma(1+\frac{1}{\beta})$	$\delta^2\Gamma(1+\frac{2}{\beta})-\mu^2$

Definition. For any $r > 0$, the *gamma function* is $\Gamma(r) = \int_0^\infty x^{r-1} e^{-x}\, dx$.

Result. $\Gamma(r) = (r-1)\Gamma(r-1)$. In particular, if r is a positive integer, then $\Gamma(r) = (r-1)!$.

Result. The exponential distribution is the only continuous memoryless distribution.
That is, $P(X > x + c \mid X > x) = P(X > c)$.

Definition. A *lifetime* distribution is continuous with range $[0, \infty)$.

Modeling lifetimes. Some useful lifetime distributions are the exponential, Erlang, gamma, and Weibull.