R. Koenker Spring 2017

Economics 574
Problem Set 1

1.: Suppose X, Y are random variables with joint density

$$f(x,y) = x^2 + xy/3 \qquad x \in [0,1], y \in [0,2].$$

Find: (a) the joint df, (b) the marginal density of X, (c) the conditional density of Y given X, (d) the conditional expectation of Y given X, (e) the unconditional expectation of Y, (f) the conditional and unconditional variance of Y given X.

2.: Suppose X_1 and X_2 are independent and uniformly distributed on $[0,1]$. Let $Y_1 = \min(X_1, X_2)$ and $Y_2 = \max(X_1, X_2)$.

(a): Show that the joint density of Y_1, Y_2 is given by

$$f(y_1, y_2) = 2I(0 < y_1 \leq y_2 < 1)$$

(b): Find the conditional density of Y_2 given Y_1. Are Y_1 and Y_2 independent?

(c): Show that

$$Esin(Y_1 Y_2 | Y_1 = y_1) = \frac{1}{y_1(1-y_1)}[cosy_1^2 - cosy_1]$$

3.: Let V be a uniformly distributed random variable on $[0, 2\pi]$ and let $X = \cos(V)$ and $Y = \sin(V)$. Show that $X \perp Y$ but $X \perp\!\!\!\perp Y$, does not hold, i.e. X and Y are not independent. Explain.

4.: [A Berlin Problem] Suppose we have uncorrelated r.v.'s $X_1, ..., X_n$, but not necessarily independent. Assume, in addition, that each X_i is symmetrically distributed around zero, i.e. $F_X(-x) = 1 - F_X(x)$ for all $x \in \Re$. Is $S_n = \sum X_i$ necessarily symmetric about zero?
Hint: Consider the example, $X \sim U[-1/2, 1/2]$, and $Y|X \sim U[-|X - 1/2|, |X - 1/2|]$.

Moments are frequently employed to characterize r.v.'s, particularly in econometrics. This is frequently useful, but occasionally dangerous. The following sequence of problems illustrate some "bad moments in statistical folklore."

5.: Suppose $Z = \log X$ is $\mathcal{N}(0,1)$, so X has the log normal distribution with density

$$f(x) = \phi(\log x)/x = \frac{1}{\sqrt{2\pi}} \frac{1}{x} e^{-(\log x)^2/2}.$$

The moments of X may be computed from the mgf of Z as

$$EX^r = m_Z(r) = \exp(\frac{1}{2}r^2).$$

(a): Why? Generalize to $\log X \sim \mathcal{N}(\mu, \sigma^2)$ and interpret.

(b): But the mgf of X is divergent for all real t suggesting that there may be other distributions with the same moment sequence. Consider the density

$$g(x) = f(x)(1 + \sin(2\pi \log x))$$

Plot f and g to get some feeling for how different they are and then show that they have the same moment sequence.

Hint: Show that

$$EX^k = \int_0^\infty x^k g(x)dx \quad since \quad \frac{1}{\sqrt{2\pi}}e^{k^2/2}\int_{-\infty}^{\infty} e^{-s^2/2} sin(2\pi s)ds = 0,$$

for $k = 0, 1, 2, ...$

6.: The Cauchy distribution with standardized density $f(x) = 1/\pi(1 + x^2)$ is the source of many interesting pathological examples.

(a): Although the Cauchy density is symmetric about 0, and rather "bell-shaped" like the normal, it fails to have a mean, i.e., EX is divergent, as are higher moments. This fact gives rise to examples of pairs of densities which "appear" rather similar, but have wildly different moment sequences. Suggest such an example based on a mixture of normal and Cauchy rv's.

(b): Using the fact that the characteristic function of a Cauchy random variable is $\phi(t) = e^{-|t|}$, show that the sample mean of n independent Cauchy random variables, $\bar{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$, follows the same Cauchy distribution.

(c): Generalize (b) to weighted averages $\tilde{X}_n = \sum_{i=1}^n w_i X_i$ with $w_i > 0$ and $\sum w_i = 1$.

7.: Lest problems (3.) and (4.) leave the impression that the inadequacy of moments to characterize distributions depends upon pathological behavior of the mgf, we give an example in which the mgf is perfectly innocuous. Let $f(x)$ be the standard normal density and

$$g(x) = f(x)(1 + 1/2\sin(2\pi x))$$

Compare f and g by plotting, and then compare their moments or cumulants.
Hint: Verify that the kgf of g is

$$k(t) = \frac{1}{2}t^2 + \log(1 + 1/2e^{-2\pi^2}\sin(2\pi t))$$

This might convey yet another wrong impression that because the mgf/kgf are almost indistinguishable the characteristic functions are also. This isn't the case: it can be shown that the characteristic functions for the normal and the McCullagh densities differ by a purely imaginary component of the form

$$d(t) = \frac{1}{4}ie^{-\frac{1}{2}(2\pi+t)^2}(e^{4\pi t} - 1)$$

which, when plotted reveals that there are significant descrepancies around ± 8. Indeed the Plancherel identity asserts that the L_2 difference in densities equals the L_2 difference in characteristic functions so this implies that big differences in densities *are* detectable by differences in characteristic functions, even though they may not be detectable by moments.

8.: Draw some moral from the proceeding three problems. You might try to work in some comment on method of moment spaces discussion at the end of Lecture 1.

9.: Suppose $Z_1, \ldots, Z_n$ are iid $\mathcal{N}(0, 1)$ r.v's. Show that $\max_i Z_i/(2\log n)^{1/2} \xrightarrow{p} 1$. Interpret this result in the context of some reasonable economic model, e.g. auctions.

Hint. Write M_n for the maximum and show

$$P(M_n \leq (2\eta \log n)^{1/2}) = [1 - P(Z_1 > (2\eta \log n)^{1/2})]^n$$

and then use the exponential inequality for normal tails.
Extended Hint:

$$P(M_n \leq a) = (1 - P(Z_1 > a))^n = \Phi(a)^n$$

using the Feller inequality

$$(\frac{1}{a} - \frac{1}{a^3})\phi(a) \le 1 - \Phi(a) \le \frac{1}{a}\phi(a)$$

we have,

$$(1 - \frac{1}{a}\phi(a))^n \le P(M_n \le a) \le (1 - (\frac{1}{a} - \frac{1}{a^3})\phi(a))^n$$

Now consider the lower bound, with $a = a_n = (2\gamma \log n)^{1/2}$...

10.: Unimodal distributions play a significant role in statistical applications: a distribution function $F : \Re \to \Re$ is unimodal about a point a if it is convex for $x < a$ and concave for $x \ge a$. The corresponding density, whose existence is ensured by the convexity/concavity, is monotone as a consequence. A nice characterization of unimodal distributions is the following.

Thm (Pestana (1980)) The r.v. Z has a unimodal distribution function (about the origin) iff $Z = XY$ for rv's X, Y such that $X \perp\!\!\!\perp Y$ and $X \sim U[0, 1]$.

This result will reappear later in the course in the context of random number generation. An interesting consequence of the result is the following classical result of Khinchine.

Thm (Khinchine) $F(x)$ is a unimodal distribution (about 0) iff it has characteristic function $\Phi(t)$ representable as

$$\Phi(t) = \frac{1}{t}\int_0^t \Psi(u)du$$

for some characteristic function Ψ.

Prove. Hint: Ψ is the characteristic function of Y.

11.: This is the Tom Rothenberg Memorial Table Method of Moments Problem, and is based on an example of Tom's. An RA is hired to measure the length and width of a rectangular seminar table. He makes a measurements of the two dimensions, but just before he is to report his results his spreadsheet program crashes and he is only able to recover n *area* measurements which had been calculated

$$A_i = L_i W_i \qquad i = 1, \dots, n$$

The problem is: can we infer anything about the length and width of the table from only the area data?

Obviously, we need a model. Suppose that the original measurements arise according to the following simple specification

$$L_i = \alpha + u_i$$
$$W_i = \beta + v_i$$

with u_i and v_i mutually independent and independent over i. Assume, further, that u_i and v_i have common distribution which is symmetric about zero, has variance σ^2, and finite fourth moment.

Under these conditions we may write

$$A_i = \alpha\beta + \alpha v_i + \beta u_i + u_i v_i$$

(1) Verify that

$$\begin{aligned} EA_i = \alpha\beta, \quad VA_i &= \sigma^2(\alpha^2 + \beta^2 + \sigma^2) \\ E(A_i - \alpha\beta)^3 &= 6\alpha\beta\sigma^4 \end{aligned}$$

where by convention $\alpha \ge \beta > 0$.

(2) Result (1.) suggests the following method of moments estimator might be used,

$$\hat{\sigma}^2 \equiv \sqrt{M_3/6M_1}$$

and we can estimate $(\alpha+\beta)^2$ by

$$S^2 \equiv M_2/\hat{\sigma}^2 - \hat{\sigma}^2 + 2M_1$$

and $(\alpha-\beta)^2$ by

$$D^2 \equiv M_2/\hat{\sigma}^2 - \hat{\sigma}^2 - 2M_1$$

Provided S^2 and D^2 are positive we may define the estimators:

$$\begin{aligned}\hat{\alpha} &= (S+D)/2\\ \hat{\beta} &= (S-D)/2.\end{aligned}$$

Supply an argument for the consistency of the estimator $\hat{\theta} = (\hat{\alpha}, \hat{\beta}, \hat{\sigma}^2)$ described above.

(3) Under somewhat stronger moment conditions, it can be shown that $\sqrt{n}(\hat{\theta}-\theta)$ is asymptotically normal. Either:
   (a) Derive the precise form of the limiting distribution of $\hat{\theta}$ and explore the effect of σ on the precision of the estimates of α and β, or
   (b) Design and conduct a small monte-carlo experiment to explore the effect in (a.).

(4) What happens if the errors are multiplicative not additive?

12.: A Volleyball problem

Consider a model of volleyball competition with old-fashioned volleyball scoring in which teams only score points when they are serving. (When the serving team loses a "rally" then the serve changes to the opposing team. In new-fashioned scoring, also known as "rally" scoring, teams score irrespectively of which team serves.)

Team Serving	Prob 1 wins	Prob 2 wins
T_1	p_1	$(1-p_1)$
T_2	$(1-p_2)$	p_2

Suppose we have offensive and defensive ratings denoted ϕ and δ from the logistic paired comparison model for teams 1 and 2. (See L20 of Ec 508.) They determine the following probabilities:

$$\begin{aligned}p_1 &= 1/(1+e^{-(\phi_1-\delta_2)})\\ p_2 &= 1/(1+e^{-(\phi_2-\delta_1)})\end{aligned}$$

Now consider the scoring transition matrix determining which team scores the next point:

State	1 wins next point	2 wins next point
T_1 wins last point	α	$(1-\alpha)$
T_2 wins last point	$(1-\beta)$	β

We can relate these transition probabilities to the p's as follows:

$$\begin{aligned}\alpha &= p_1 + (1-p_1)(1-\beta)\\ \beta &= p_2 + (1-p_2)(1-\alpha)\end{aligned}$$

To see these relationships consider team 1's situation: if team 1 won the previous point then it is serving and the probability is p_1 that it gets the next point. On the other hand, if it loses its serve (which happens with probability $(1-p_1)$ then it wins the next point with probability

$(1-\beta)$. This yields the expression for α. The same reasoning for team 2 yields the expression for β.

(1) Solve for α and β in terms of p_1, p_2.

(2) Now suppose we would like to compute the probability $P_n(j)$ that team 1 wins by a score of n to j, $j < n$, assuming that team 1 serves first, which can be expressed as,

$$P_n(j) = \sum_k P(E_k)$$

where $k = (k_1, \dots, k_n)$ and E_k is the event,

$$\begin{aligned} E_k = \{ \; & T_2 \text{ scores } k_1 \text{ points, } T_1 \text{ scores a point,} \\ & T_2 \text{ scores } k_2 \text{ points, } T_1 \text{ scores a point,} \\ & \;\vdots \\ & T_2 \text{ scores } k_n \text{ points, } T_1 \text{ scores a point, } \} \end{aligned}$$

so the sum is over all n-tuples k such that $\sum_{i=1}^n k_i = j$. From our transition matrix, we have, setting

$$\begin{aligned} c_0 &= \alpha \\ c_{k_i} &= (1-\alpha)\beta^{k_i-1}(1-\beta) \qquad k_i > 0 \end{aligned}$$

that this probability is

$$P(E_k) = \prod_{i=1}^n c_{k_i}.$$

Let $\{X_i = 1, \dots, n\}$ be iid random variables with

$$P(X_i = k) = \begin{cases} \alpha & k = 0 \\ (1-\alpha)\beta^{k-1}(1-\beta) & k > 0 \end{cases}$$

and set $S_n = \sum X_i$, then

$$P_n(j) = P(S_n = j)$$

show that the generating function

$$g_{X_1}(s) = \sum_{i=1}^{\infty} c_i s^i = \alpha + \frac{(1-\alpha)(1-\beta)s}{1-\beta s}$$

(3) Show that the generating function for S_n is

$$g_{S_n}(s) = [\alpha + (1-\alpha-\beta)s]^n [1-\beta s]^{-n}$$

so to find probabilities we just need the coefficients of the power series expansion,

$$g_{S_n}(s) = \sum_{k=0}^{\infty} P(S_n = k) s^k$$

Let $\gamma \equiv (1-\alpha-\beta)$ and $\delta = \gamma/\alpha$ and write

$$((\alpha - \gamma s)/(1-\beta s))^n = \alpha^n ((1+\delta s)/(1-\beta s))^n.$$

The binomial theorem deals nicely with the numerator,

$$(1+s)^n = \sum_{k=0}^n \binom{n}{k} s^k,$$

and a somewhat more exotic form that goes back to Newton (perhaps) is

$$(1-s)^{-n} = \sum_{k=0}^{\infty} \binom{n+k-1}{k} s^k,$$

so generalizing slightly we have

$$P_n(j) = \sum_{k=0}^{j} \binom{n}{j-k} \binom{n+k-1}{k} \beta^k \alpha^{n-j+k} (1-\alpha-\beta)^{j-k}.$$

Verify and evaluate for $p_1 = .45, p_2 = .40$ and $n = 15$, the probability of various scores $j = 0, 1, \ldots, 13$. Use Mathematica or R or some other convenient computational aid.

[Historical Note] When $j > 13$ then things get complicated by the end game rules. In the ancient form of volleyball scoring teams needed to win by two points, but new forms of scoring try to simplify this. One objective of the foregoing fun and games is that it enables us to compute the length of matches for various serving rules. (This problem is based on some preliminary work that I did on a paper with Wally Hendricks that was never finished tentatively called "Redesigning Volleyball for TV." It was intended to compare various proposals to change the rules to make matches more predictable in length.) The "paired comparison" model discussed in 508 can be used to estimate ratings parameters that can then be used to evaluate the underlying probabilities in this model, and therefore to make predictions or seedings of tournaments.