Selected Problems from Pitman's "Probability" Text

Statistics 200A, Nathan Ross, Fall 2009

6.2.13 Let $S_n = X_1 + \cdots + X_n$ be the number of successes in of n independent Bernoulli (p) trials $X_1, X_2, \ldots, X_n$.

1. For $1 \leq m \leq n$, show that the conditional distribution of S_m, the number of successes in the first m trials, given $S_n = k$, is identical to the distribution of the number of good elements in a random sample of size m without replacement from a population of k good and $n - k$ bad elements.
2. Use the result of a) to rederive the result of Example 4 that $E(S_m \mid S_n = k) = mk/n$.
3. Find $Var(S_m \mid S_n = k)$.

6.2.15 Let Π be a random proportion between 0 and 1, for example, the proportion of black balls in an urn picked at random from some population of urns. Let S be the number of successes in n Bernoulli trials, which given $\Pi = p$ are independent with probability p, for example, the number of black balls in n draws at random, with replacement from the urn picked at random.

1. Find a formula for $E(S)$ in terms of n and $E(\Pi)$.
2. Find a formula for $Var(S)$ in terms of n, $E(\Pi)$, and $Var(\Pi)$.
3. For given n and $E(\Pi) = p$, say, which distribution of Π makes $Var(S)$ as large as possible? Which as small as possible? Prove your answers using your answer to b).

6.2.17 Suppose you want to predict the value of a random variable Y. Instead of just trying to predict the value of Y by a constant (such as the mean or median), suppose that some additional information pertinent to the prediction of Y is available. For instance, you might know the value of some other random variable X, whose joint distribution with Y is assumed known. The problem here is to predict the value of Y by a function of X, call it $g(X)$. Once the value x of X is known, the value $g(x)$ of $g(X)$ can be calculated and used to predict the unknown value of Y.

One measure of the goodness of the predictor $g(X)$ is its *mean square error* (MSE)

$$MSE(g(X)) = E[(Y - g(X))^2]$$

It is a measure of, on average, how far off the prediction is. Show that $g(X) = E(Y|X)$ minimizes the MSE. *Hint: Condition on the value of* X

$$E[(Y - g(X))^2] = \sum_x E[(Y - g(X))^2 \mid X = x]P(X = x)$$

and minimize each term in the sum separately.

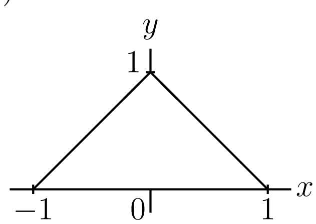
6.3.2 Let X and Y have the following joint density:

$$f(x,y) = \begin{cases} 2x + 2y - 4xy & \text{for } 0 \le x \le 1 \text{ and } 0 \le y \le 1 \\ 0 & \text{otherwise} \end{cases}$$

1. Find the marginal densities of X and Y.
2. Find $f_Y(y \mid X = \frac{1}{4})$.
3. Find $E(Y \mid X = \frac{1}{4})$.

6.3.5 Suppose (X, Y) has uniform distribution on the triangle shown in the diagram. For x between -1 and 1, find:

1. $P(Y \ge \frac{1}{2} | X = x)$;
2. $P(Y < \frac{1}{2} | X = x)$;
3. $E(Y | X = x)$;
4. $Var(Y | X = x)$.



6.3.6 Suppose X, Y are random variables with joint density

$$f_{X,Y}(x,y) = \frac{1}{2\pi\sqrt{x(y-x)}} e^{-y/2} \qquad (0 < x < y)$$

1. Find the distribution of Y. *Hint: For integration use the substitution $x = ys$.*
2. Compute $E(X | Y = 1)$.

6.3.8 The random variable X has a uniform distribution on $(0, 1)$. Given that $X = x$, the random variable Y is binomial with parameters $n = 5$ and $p = x$.

1. Find $E(Y)$ and $E(Y^2)$.
2. Find $P(Y = y \text{ and } x < X < x + dx)$.
3. Find the density of X given $Y = y$. Do you recognize it? If yes, as what?

6.3.13 Let X and Y be independent random variables, X with uniform distribution on $(0, 3)$, Y with Poisson (λ) distribution. Find:

1. a formula in terms of λ for $P(X < Y)$;
2. the conditional density of X given $X < Y$, and sketch its graph in the cases $\lambda = 1, 2, 3$;

3. $E(X|X < Y)$.

6.4.22 Suppose there were m married couples, but that d of these $2m$ people have died. Regard the d deaths as striking the $2m$ people at random. Let X be the number of surviving couples. Find:

1. $E(X)$;
2. $Var(X)$.

6.4.23 For random variables X and Y, the *linear prediction problem* for predicting Y based on knowledge of X is the problem of finding a linear function of X, $\beta X + \gamma$, which minimizes the *mean square* of the prediction error

$$MSE = E[Y - (\beta X + \gamma)]^2$$

(Compare with Exercise 6.2.17 where the predictor of Y could be an arbitrary function of X.) This exercise derives the basic formulae for the best linear predictor according to this criterion.

1. Expand out the MSE using algebra, and regard it as a quadratic function of γ and β with coefficients involving the numbers $E(X)$, $E(Y)$, $E(XY)$, etc.
2. Differentiate this function with respect to γ to show that for fixed β, the unique γ which minimizes the MSE is $\hat{\gamma}(\beta) = E(Y) - \beta E(X)$. What is the resulting minimal MSE called when $\beta = 0$?
3. Consider now the MSE as a function of β, with $\gamma = \hat{\gamma}(\beta)$ the best γ for the given β. Differentiate this function with respect to β, and show that it is minimized at $\hat{\beta} = Cov(X, Y)/Var(X)$ where it is assumed that $Var(X) > 0$.
4. Deduce that the unique pair (β, γ) which minimizes the MSE is $(\hat{\beta}, \hat{\gamma}(\hat{\beta}))$.
5. Let $\hat{Y} = \hat{\beta}X + \hat{\gamma}$ now denote this best linear predictor. Show that

$$E(\hat{Y}) = E(Y); \quad Var(\hat{Y}) = \hat{\beta}^2 Var(X); \quad E[\hat{Y}(Y - \hat{Y})] = 0$$

6. Deduce that the variance of Y can be decomposed into the sum of the variance of the best predictor $\hat{Y}$ and the minimum MSE according to the formula

$$Var(Y) = Var(\hat{Y}) + E[(Y - \hat{Y})^2]$$

with $Var(\hat{Y}) = \rho^2 Var(Y)$ and $E[(Y - \hat{Y})^2] = (1 - \rho^2) Var(Y)$ where $\rho = Corr(X, Y)$.

7. It is customary to express the slope $\hat{\beta}$ of the best linear predictor $\hat{Y} = \hat{\beta}X + \hat{\gamma}$ in terms of ρ. Show that $\hat{\beta} = \rho SD(Y)/SD(X)$ and that the intercept $\hat{\gamma}$ is then uniquely determined by the requirement that the line $y = \hat{\beta}x + \hat{\gamma}$ passes through the point $(E(X), E(Y))$.

8. Let $Y^* = (Y - E(Y))/SD(Y)$, $X^* = (X - E(X))/SD(X)$. Show that the best linear predictor of Y^* based on X^* is just ρX^*. So the correlation coefficient ρ is simply the slope of the best linear predictor when the variables are expressed in standard units.

6.5.7 Let X and Y have bivariate normal distribution with parameters μ_X, μ_Y, ${\sigma_X}^2$, ${\sigma_Y}^2$, and ρ. Show that X and Y are independent if and only if they are uncorrelated.

6.5.9 Suppose that W has normal (μ, σ^2) distribution. Given that $W = w$, suppose that Z has normal $(a\,w + b, \tau^2)$ distribution.

1. Show the joint distribution of W and Z is bivariate normal, and find its parameters.
2. What is the distribution of Z?
3. What is the conditional distribution of W given $Z = z$?

6.5.10 Show that if V and W have a bivariate normal distribution then

1. every linear combination $aV + bW$ has a normal distribution;
2. every pair of linear combinations $(aV + bW, cV + dW)$ has a bivariate normal distribution.
3. Find the parameters of the distributions obtained in a) and b) in terms of the parameters of the joint distribution of V and W.

6.5.11 Show that for standard bivariate normal variables X and Y with correlation ρ,

$$E(\max(X, Y)) = \sqrt{\frac{1-\rho}{\pi}}$$