Stat 5101 Notes: Algorithms (thru 2nd midterm)

Charles J. Geyer

October 18, 2012

Contents

1 Calculating an Expectation or a Probability

Probability is a special case of expectation (deck 1, slide 62).

1.1 From a PMF

If f is a PMF having domain S (the *sample space*) and g is any function, then

$$E\{g(X)\} = \sum_{x \in S} g(x) f(x)$$

(deck 1, slide 56 and deck 3, slide 12), and for any event A (a subset of S)

$$\begin{aligned} \Pr(A) &= \sum_{x \in S} I_A(x) f(x) \\ &= \sum_{x \in A} f(x) \end{aligned}$$

(deck 1, slide 62).

1.2 From a PDF

If f is a PDF having domain S (the *sample space*) and g is any function, then

$$E\{g(X)\} = \int_S g(x) f(x) \, dx$$

(deck 3, slide 66), and for any event A (a subset of S)

$$\begin{aligned} \Pr(A) &= \int_S I_A(x) f(x) \, dx \\ &= \int_A f(x) \, dx \end{aligned}$$

(deck 1, slide 84).

And similarly for random vectors,

$$E\{g(\mathbf{X})\} = \int_S g(\mathbf{x}) f(\mathbf{x}) \, d\mathbf{x}$$

(deck 3, slide 66), and

$$\begin{aligned} \Pr(A) &= \int_S I_A(\mathbf{x}) f(\mathbf{x}) \, d\mathbf{x} \\ &= \int_A f(\mathbf{x}) \, d\mathbf{x} \end{aligned}$$

(deck 1, slide 84), where the boldface indicates vectors and the integral signs indicate multiple integrals (the same dimension as the dimension of $\mathbf{x}$).

1.3 From given Expectations using Uncorrelated

If X and Y are uncorrelated random variables, then

$$E(XY) = E(X)E(Y)$$

(deck 2, slide 73).

1.4 From given Expectations using Independent

If X and Y are independent random variables and g and h are any functions, then

$$E\{g(X)h(Y)\} = E\{g(X)\}E\{h(Y)\}$$

(deck 2, slide 76).

More generally, if X_1, X_2, ..., X_n are independent random variables and g_1, g_2, ..., g_n are any functions, then

$$E\left\{\prod_{i=1}^{n} g_i(X_i)\right\} = \prod_{i=1}^{n} E\{g_i(X_i)\}$$

(deck 2, slide 76).

1.5 From given Probabilities using Inclusion-Exclusion

$$\begin{aligned}
\Pr(A \cup B) &= \Pr(A) + \Pr(B) - \Pr(A \cap B) \\
\Pr(A \cup B \cup C) &= \Pr(A) + \Pr(B \cup C) - \Pr(A \cap (B \cup C)) \\
&= \Pr(A) + \Pr(B \cup C) - \Pr((A \cap B) \cup (A \cap C)) \\
&= \Pr(A) + \Pr(B) + \Pr(C) - \Pr(B \cap C) \\
&\qquad - \Pr(A \cap B) - \Pr(A \cap C) + \Pr(A \cap B \cap C)
\end{aligned}$$

and so forth (deck 2, slides 139 and 140).

In the special case that A_1, ..., A_n are mutually exclusive events

$$\Pr\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} \Pr(A_i)$$

(deck 2, slide 137).

1.6 From given Probabilities using Complement Rule

$$\Pr(A^c) = 1 - \Pr(A)$$

(deck 2, slide 143).

1.7 From given Probabilities using Independent

In the special case that A_1, ..., A_n are independent events

$$\Pr\left(\bigcap_{i=1}^{n} A_i\right) = \prod_{i=1}^{n} \Pr(A_i)$$

(deck 2, slide 147).

1.8 From given Expectations using Linearity of Expectation

1.8.1 Expectation of Sum and Average

If X_1, X_2, ..., X_n are random variables, then

$$E\left\{\sum_{i=1}^{n} X_i\right\} = \sum_{i=1}^{n} E(X_i)$$

(deck 2, slide 10). In particular, if X_1, X_2, ..., X_n all have the same expectation μ, then

$$E\left\{\sum_{i=1}^{n} X_i\right\} = n\mu$$

(deck 2, slide 90), and

$$E(\overline{X}_n) = \mu,$$

where

$$\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i \tag{1}$$

(deck 2, slide 90).

1.8.2 Variance of Sum and Average

If X_1, X_2, ..., X_n are random variables, then

$$\begin{aligned}
\operatorname{var}\left\{\sum_{i=1}^{n} X_i\right\} &= \sum_{i=1}^{n}\sum_{j=1}^{n} \operatorname{cov}(X_i, X_j) \\
&= \sum_{i=1}^{n} \operatorname{var}(X_i) + \sum_{i=1}^{n}\sum_{j=1 j\neq i}^{n} \operatorname{cov}(X_i, X_j) \\
&= \sum_{i=1}^{n} \operatorname{var}(X_i) + 2\sum_{i=1}^{n}\sum_{j=i+1}^{n} \operatorname{cov}(X_i, X_j)
\end{aligned}$$

(deck 2, slide 71). In particular, if X_1, X_2, ..., X_n are uncorrelated, then

$$\operatorname{var}\left\{\sum_{i=1}^{n} X_i\right\} = \sum_{i=1}^{n} \operatorname{var}(X_i)$$

(deck 2, slide 75). More particularly, if X_1, X_2, $\ldots$, X_n are uncorrelated and all have the same variance σ^2, then

$$\operatorname{var}\left\{ \sum_{i=1}^{n} X_i \right\} = n\sigma^2$$

(deck 2, slide 90) and

$$\operatorname{var}(\overline{X}_n) = \frac{\sigma^2}{n}$$

(deck 2, slide 90), where $\overline{X}_n$ is given by (1).

1.8.3 Expectation and Variance of Linear Transformation

If X is a random variable and a and b are constants, then

$$\begin{aligned} E(a + bX) &= a + bE(X) \\ \operatorname{var}(a + bX) &= b^2 \operatorname{var}(X) \end{aligned}$$

(deck 2, slide 8 and slide 37).

1.8.4 Covariance of Linear Transformations

If X and Y are random variables and a, b, c, and d are constants, then

$$\operatorname{cov}(a + bX, c + dY) = bd \operatorname{cov}(X, Y)$$

(homework problem 3-7).

1.8.5 Expectation and Variance of Vector Linear Transformation

If $\mathbf{X}$ is a random vector, $\mathbf{a}$ is a constant vector, and $\mathbf{B}$ is a constant matrix such that $\mathbf{a} + \mathbf{B}\mathbf{X}$ makes sense (the dimension of $\mathbf{a}$ and the row dimension of $\mathbf{B}$ are the same, and the dimension of $\mathbf{X}$ and the column dimension of $\mathbf{B}$ are the same), then

$$\begin{aligned} E(\mathbf{a} + \mathbf{B}\mathbf{X}) &= \mathbf{a} + \mathbf{B}E(\mathbf{X}) \\ \operatorname{var}(\mathbf{a} + \mathbf{B}\mathbf{X}) &= \mathbf{B} \operatorname{var}(\mathbf{X}) \mathbf{B}^T \end{aligned}$$

(deck 2, slide 64).

1.8.6 "Short Cut" Formulas

If X is a random variable, then

$$\operatorname{var}(X) = E(X^2) - E(X)^2$$

(deck 2, slide 21).

If X and Y are random variables, then

$$\operatorname{cov}(X, Y) = E(XY) - E(X)E(Y)$$

(homework problem 3-7).

1.9 From Moment Generating Function

$$E(X^k) = \varphi^{(k)}(0)$$

(deck 3, slide 20), where φ is the moment generating function of the random variable X and the right-hand side denotes the k-th derivative of φ evaluated at zero.

1.10 Convolution Formula

If X and Y are independent random variables in the same probability model and $Z = X + Y$ and f_X, f_Y, and f_Z denote the PMFs of these random variables, then

$$f_Z(z) = \sum_x f_X(x) f_Y(z - x)$$

(deck 3, slide 52), where the sum ranges over the possible values of X and f_Y is extended to be zero off of the support of Y.

2 Change of Variable Formulas

2.1 Discrete Distributions

2.1.1 One-to-one Transformations

If f_X is the PMF of the random variable X, if $Y = g(X)$, and g is an invertible function with inverse function h (that is, $X = h(Y)$), then

$$f_Y(y) = f_X[h(y)]$$

and the domain of f_Y is the range of the function g (the set of possible Y values) (deck 2, slide 86).

2.1.2 Many-to-one Transformations

If f_X is the PMF of the random variable X having sample space S and $Y = g(X)$, then

$$f_Y(y) = \sum_{\substack{x \in S \\ g(x)=y}} f_X(x)$$

and the domain of f_Y is the codomain of the function g (deck 2, slide 81).

2.2 Continuous Distributions

2.2.1 One-to-one Transformations

Univariate If f_X is the PDF of the random variable X, if $Y = g(X)$, and g is an invertible function with inverse function h (that is, $X = h(Y)$), then

$$f_Y(y) = f_X[h(y)] \cdot |h'(y)|$$

and the domain of f_Y is the range of the function g (the set of possible Y values) (deck 3, slide 123).

Linear Transformation (Special Case of Above) If f_X is the PDF of the random variable X defined on the whole real line and $Y = a + bX$ with $b > 0$, then the PDF of Y is

$$f_Y(y) = \frac{1}{b} f_X\left(\frac{y-a}{b}\right)$$

and the domain of f_Y is the whole real line (deck 3, slide 140).

Multivariate If $f_{\mathbf{X}}$ is the PDF of the random vector $\mathbf{X}$, if $\mathbf{Y} = g(\mathbf{X})$, and g is an invertible function with inverse function h (that is, $\mathbf{X} = h(\mathbf{Y})$), then

$$f_{\mathbf{Y}}(\mathbf{y}) = f_{\mathbf{X}}[h(\mathbf{y})] \cdot |\det J(\mathbf{y})|$$

and the domain of $f_{\mathbf{Y}}$ is the range of the function g (the set of possible $\mathbf{Y}$ values) (deck 3, slide 122), where $J(\mathbf{y})$ is the Jacobian matrix of the transformation h, the matrix whose i, j component is $\partial h_i(\mathbf{y}) / \partial y_j$, where

$$h(\mathbf{y}) = \big(h_1(\mathbf{y}), \ldots, h_n(\mathbf{y})\big)$$

(deck 3, slide 121).

2.2.2 Alternate Method (Does not Require One-to-one)

If f_X is the PDF of the random variable X and $Y = g(X)$, then the DF of Y is

$$F_Y(y) = \Pr(Y \leq y) = \Pr\{g(X) \leq y\}$$

(deck 3, slide 154), and then the PDF of Y can be found by differentiation (Section 4.1 below).

3 PMF or PDF and Independence

If X_1, ..., X_n are independent random variables having PMF or PDF f_1, ..., f_n, then

$$f(x_1, \ldots, x_n) = \prod_{i=1}^{n} f_i(x_i)$$

(deck 1, slide 98 and deck 3, slide 115). In this terminology of deck 5 (Section 6 below) this says the joint is the product of the marginals (when the components are independent).

In particular, if X_1, ..., X_n are independent and identically distributed random variables having PMF or PDF h, then

$$f(x_1, \ldots, x_n) = \prod_{i=1}^{n} h(x_i).$$

4 PDF or PMF to DF and Vice Versa

4.1 DF to PDF

If F is the DF of a continuous random variable X, and f is defined by

$$f(x) = F'(x), \qquad \text{at points } x \text{ where } F \text{ is differentiable}$$

(and may be arbitrarily defined at other points), then f is a PDF of X (deck 3, slide 105) defined on the whole real line.

4.2 PDF to DF

If f is the PDF of a continuous random variable defined on the whole real line and F is the corresponding DF, then

$$\begin{aligned} F(x) &= \Pr(X \le x) \\ &= \int_{-\infty}^{x} f(s)\, ds \end{aligned}$$

(deck 3, slides 92 and 104).

4.3 PDF to DF (Alternative Method)

If f is the PDF of a random variable defined by case splitting, say

$$f(x) = \begin{cases} f_1(x), & x < a_1 \\ f_2(x), & a_1 < x < a_2 \\ f_3(x), & a_2 < x < a_3 \\ f_4(x), & a_3 < x \end{cases}$$

then the DF is

$$F(x) = \int f_i(x)\, dx$$

on each of the four intervals, the indefinite integral containing an arbitrary constant, which is different for each interval, and the constants must be adjusted to make F have the properties of a DF given on deck 3, slides 112 and 113, that is F is continuous, goes to zero as $x \to -\infty$ and goes to one as $x \to +\infty$.

4.4 DF to PMF

If f is the PMF of a discrete random variable defined on the whole real line and F is the corresponding DF, then $f(x)$ is the size of the jump F makes at x, that is,

$$f(x) = F(x) - \sup_{y<x} F(y)$$

(deck 3, slide 109). In particular, $f(x) = 0$ whenever F is continuous at x.

4.5 PMF to DF

If $f : S \to \mathbb{R}$ is the PMF of a discrete random variable and F is the corresponding DF, then

$$\begin{aligned} F(x) &= \Pr(X \leq x) \\ &= \sum_{\substack{s \in S \\ s \leq x}} f(s) \end{aligned}$$

(deck 3, slide 109).

5 Quantiles

5.1 Continuous Random Variables

If F is the DF of a continuous random variable, then the p-th quantile of this random variable or this distribution is any x satisfying

$$F(x) = p$$

and the solution is unique if F is differentiable at x with $F'(x) > 0$ (deck 4, slide 2).

The solution is non-unique only if F has a flat section with value p, that is if there exist points a and b with $a < b$ such that

$$F(x) = p, \qquad a \leq x \leq b$$

in which case every $x \in [a, b]$ is a p-th quantile.

5.2 Discrete Random Variables

If F is the DF of a discrete random variable, then the p-th quantile of this random variable or this distribution is any x satisfying

$$F(y) \leq p \leq F(x), \qquad \text{for all } y < x$$

(deck 4, slide 2).

There are two cases. The solution is unique if we do not have $F(x) = p$ for any x. In that case the p-th quantile is the point x where F jumps past p, that is,

$$F(y) < p < F(x), \qquad y < x.$$

The solution is non-unique if we do have $F(x) = p$ for some x, in which case (because the DF of a discrete random variable is a step function) we must have $F(x) = p$ for all x in an interval, and every such x is a p-th quantile.

6 Joint, Marginal, Conditional

The probability distribution of a random vector is often called the *joint* distribution of the components of the random vector.

In this context the probability distribution of a random vector comprising some subset of these components is called a *marginal* distribution.

The terms *joint* and *marginal* have meaning only in context. They have no context-independent meaning.

We can say $f(x, y, z)$ is the joint PDF of X, Y, and Z, and $f(x, y)$, $f(y, z)$, and $f(x, z)$ are three marginal PDF. So in one context $f(x, y, z)$ is a joint distribution and $f(x, y)$ is a marginal. But in another context we can say $f(x, y)$ is a joint distribution and $f(x)$ and $f(y)$ are its marginals. (Pedantically, each f here should have a different letter or distinguishing subscript because each is a different function.)

6.1 Joint to Marginal

To derive a marginal PMF or PDF from a joint PMF or PDF, sum or integrate out the variable(s) you don't want (sum if discrete, integrate if continuous) (deck 5, slides 4–9). Given the joint of X and Y, to obtain the marginal of X sum out y

$$f_X(x) = \sum_y f_{X,Y}(x, y)$$

or integrate out y

$$f_X(x) = \int f_{X,Y}(x, y)\, dy$$

as the case may be.

Given the joint of X, Y, and Z, to obtain the marginal of X sum out y and z

$$f_X(x) = \sum_y \sum_z f_{X,Y,Z}(x, y, z)$$

or integrate out y and z

$$f_X(x) = \iint f_{X,Y,Z}(x, y, z)\, dy\, dz$$

as the case may be.

Given the joint of X, Y, and Z, to obtain the (bivariate) marginal of X and Y sum out z

$$f_{X,Y}(x, y) = \sum_z f_{X,Y,Z}(x, y, z)$$

or integrate out z

$$f_{X,Y}(x, y) = \int f_{X,Y,Z}(x, y, z)\, dz$$

as the case may be.

And so forth (many special cases of the general principle could be given).

6.2 Joint to Conditional

$$\text{conditional} = \frac{\text{joint}}{\text{marginal}}$$

and the marginal is the marginal for the variable(s) behind the bar in the conditional

$$\begin{aligned} f(x \mid y) &= \frac{f(x, y)}{f(y)} \\ f(w, x \mid y, z) &= \frac{f(w, x, y, z)}{f(y, z)} \end{aligned}$$

and so forth (pedantically, each f here should have a different letter or distinguishing subscript because each is a different function).

6.3 Conditional and Marginal to Joint

$$\text{joint} = \text{conditional} \times \text{marginal}$$

and the marginal is the marginal for the variable(s) behind the bar in the conditional

$$\begin{aligned} f(x, y) &= f(x \mid y) f(y) \\ f(w, x, y, z) &= f(w, x \mid y, z) f(y, z) \end{aligned}$$

and so forth (pedantically, each f here should have a different letter or distinguishing subscript because each is a different function).

Note that the equations in this section are the equations in the preceding section rearranged.

6.4 Conditional Probability and Expectation

Everything in Sections 1 to 5 above can be redone for conditional probability distributions. The variables behind the bar in the conditional distribution just go along for the ride like parameters in unconditional distributions (deck 5, slides 11–22). Calculating expectations (compare with Section 1 above)

$$E\{g(X) \mid Y\} = \int g(x) f(x \mid y) \, dx$$
$$E\{g(W, X) \mid Y, Z\} = \int g(w, x) f(w, x \mid y, z) \, dw \, dx$$

(replace integrals with sums if W and X are discrete). Calculating probabilities (compare with Section 1 above)

$$\Pr\{X \in A \mid Y\} = \int_A f(x \mid y) \, dx$$
$$\Pr\{(W, X) \in A \mid Y, Z\} = \int I_A(w, x) f(w, x \mid y, z) \, dw \, dx$$

(replace integrals with sums if W and X are discrete). PDF to DF (compare with Section 4.2 above)

$$F(x \mid y) = \int_{-\infty}^{x} f_{X|Y}(s \mid y)$$

(replace the integral with a sum if X is discrete). Quantiles (compare with Section 5 above, if F is the conditional DF of X given Y and X is continuous, then the p-th quantile of this distribution is the x that satisfies

$$F(x \mid y) = p.$$