12-2017

Probability Models and Compounding

David L. Farnsworth
Rochester Institute of Technology

James E. Marengo
Rochester Institute of Technology

Recommended Citation

Farnsworth, David L. and Marengo, James E., "Probability Models and Compounding" (2017). *The Mathematical Scientist*, 42 (3), 117-122. Accessed from
https://repository.rit.edu/article/2121

Math. Scientist **42,** 117–122 (2017)
Printed in England

PROBABILITY MODELS AND COMPOUNDING

JAMES E. MARENGO * ** AND
DAVID L. FARNSWORTH,* *** *Rochester Institute of Technology*

Abstract

We present the case that the ideas contained in a particular sequence of formulas are important in probability and statistics. The synthesis offered by the concepts in the sequence can be very valuable. Facility with this sequence and its underpinnings should be in the skill set of anyone who uses or studies probability or statistics. For illustrative purposes, we give applications to mixture distributions and Bayesian analyses.

Keywords: Probability model; conditional probability; Bayesian analysis; law of total probability; mixture distribution; compounding distribution

2010 Mathematics Subject Classification: Primary 62E15
Secondary 00A71; 62C10

1. Introduction

We investigate a set of ideas that applies to both probability and statistics at all levels. The concepts that are used include conditional probability, the law of total probability, marginal probability distribution, compound distribution, weighted average, expected value, Bayes' theorem, and a mixture of random variables or of probability distributions.

A synthesis is obtained from a sequence of five formulas. The sequence may be used in whole or in part depending upon the situation. In broad terms, it is summarized as follows. We know the conditional distribution,

$$f_{X \mid Y}(x; \lambda \mid y), \tag{1}$$

of X given Y, where the parameter λ can be a vector. Also, we know the distribution of Y,

$$f_Y(y; \theta), \tag{2}$$

where the parameter θ can be a vector. The joint distribution of X and Y is created by the product of (1) and (2):

$$f_{X,Y}(x, y; \lambda, \theta) = f_{X \mid Y}(x; \lambda \mid y) f_Y(y; \theta). \tag{3}$$

The unconditional distribution of X is the marginal distribution

$$f_X(x; \lambda, \theta) = \int_{-\infty}^{\infty} f_{X,Y}(x, y; \lambda, \theta)\, \mathrm{d}y. \tag{4}$$

The conditional distribution of Y given X is the ratio of (3) and (4):

$$f_{Y \mid X}(y; \lambda, \theta \mid x) = \frac{f_{X,Y}(x, y; \lambda, \theta)}{f_X(x; \lambda, \theta)}. \tag{5}$$

Received 2 June 2017; revision received 12 June 2017.
* Postal address: School of Mathematical Sciences, Rochester Institute of Technology, Rochester, NY 14623, USA.
** Email address: jemsma@rit.edu
*** Email address: dlfsma@rit.edu

The probability distributions may be discrete, continuous, or partly discrete and partly continuous. For discrete distributions, summation replaces integration.

Section 2 contains two examples in order to set the notation for the applications in Section 3. To show the diverse variety of settings in which (1)–(5) is relevant, we discuss applications to mixture models and Bayesian analysis in Section 3. Closing comments are in Section 4.

2. Examples

We present two specific examples of $f_{X\mid Y}(x;\lambda \mid y)$ and $f_Y(y;\theta)$ in order to make the ideas more concrete and to introduce the notation for the applications in Section 3.

Example 1. The distribution of X given $Y = P$ in (1), where λ is the single parameter n, is the conditional binomial probability mass function

$$f_{X\mid P}(x;n \mid p) = \binom{n}{x} p^x (1-p)^{n-x}, \tag{6}$$

for $x \in \{0,1,2,\dots,n\}$ and $p \in (0,1)$. The parameter n is a positive integer and is interpreted as the sample size. For a given value p of the probability, the expected value of X is $\mathbb{E}(X \mid P = p) = \mathbb{E}(X \mid p) = np$; see [3, p. 55] and [4, pp. 108–113].

We take the distribution of $Y = P$ in (2) to be the beta distribution, i.e.

$$f_P(p;\alpha,\beta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} p^{\alpha-1}(1-p)^{\beta-1}, \tag{7}$$

for $p \in (0,1)$. The parameter θ is the vector (α,β) , where $\alpha > 0$ and $\beta > 0$. The parameters can be selected according to our knowledge about P; see [2, pp. 145–150]. Distribution (7) is the weighting function for the possible values of P in (6). The expected value of P is

$$\mathbb{E}(P) = \frac{\alpha}{\alpha+\beta}; \tag{8}$$

see [3, pp. 107–109] and [4, pp. 167–168]. Since the total probability is 1, we have the integration formula

$$\int_0^1 p^{\alpha-1}(1-p)^{\beta-1}\,\mathrm{d}p = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}. \tag{9}$$

The product of the distributions (6) and (7) is the joint distribution according to (3),

$$f_{X,P}(x,p;n,\alpha,\beta) = \binom{n}{x}\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} p^{x+\alpha-1}(1-p)^{n+\beta-x-1}, \tag{10}$$

for $x \in \{0,1,2,\dots,n\}$ and $p \in (0,1)$, which is a mixed discrete and continuous distribution.

By integrating (10) over all values of p as in (4), a weighted average is obtained. This is an application of the law of total probability. The unconditional distribution of X is a sum, which is expressed as an integral, over each and every possible value of P. That marginal distribution is

$$\begin{aligned} f_X(x;n,\alpha,\beta) &= \int_0^1 \binom{n}{x}\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} p^{x+\alpha-1}(1-p)^{n+\beta-x-1}\,\mathrm{d}p \\ &= \binom{n}{x}\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \int_0^1 p^{x+\alpha-1}(1-p)^{n+\beta-x-1}\,\mathrm{d}p \\ &= \binom{n}{x}\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\,\frac{\Gamma(x+\alpha)\Gamma(n+\beta-1)}{\Gamma(n+\alpha+\beta)}, \end{aligned} \tag{11}$$

for $x \in \{0, 1, 2, \ldots, n\}$, where the second integral is evaluated using (9). This distribution can be viewed as an average of distribution (6) over values of P for each X, since it can be written as the expected value of (6),

$$\begin{aligned} f_X(x; n, \alpha, \beta) &= \mathbb{E}\left(\binom{n}{x} p^x (1-p)^{n-x} \,\middle|\, x\right) \\ &= \int_0^1 \left[\binom{n}{x} p^x (1-p)^{n-x}\right] \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} p^{\alpha-1}(1-p)^{\beta-1}\,\mathrm{d}p. \end{aligned}$$

The general formula in (4) and the distribution (11) are called *compound distributions*; see [5, p. 191].

Using (5), the distribution of P is obtained for each value of X as

$$\begin{aligned} &f_{P\,|\,X}(p; n, \alpha, \beta \mid x) \\ &\quad = \binom{n}{x} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} p^{x+\alpha-1}(1-p)^{n+\beta-x-1} \Big/ \binom{n}{x} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \frac{\Gamma(x+\alpha)\Gamma(n+\beta-1)}{\Gamma(n+\alpha+\beta)} \\ &\quad = \frac{\Gamma(n+\alpha+\beta)}{\Gamma(x+\alpha)\Gamma(n+\beta-x)} p^{x+\alpha-1}(1-p)^{n+\beta-x-1}, \end{aligned} \tag{12}$$

for $p \in (0, 1)$, which is a beta distribution. This formula might be used to estimate P subsequent to an experimental outcome $X = x$, based on employing the input functions (6) and (7). Following (8), the expected value of P is

$$\mathbb{E}(P \mid X = x) = \mathbb{E}(P \mid x) = \frac{x+\alpha}{(x+\alpha)+(n+\beta-x)} = \frac{x+\alpha}{n+\alpha+\beta}. \tag{13}$$

It is interesting that the expected value (13) is a weighted average of an estimated value of P from a sample x/n and the mean $\alpha/(\alpha+\beta)$ in (8) of the original beta distribution for P, since

$$\mathbb{E}(P \mid x) = \frac{n}{n+\alpha+\beta}\,\frac{x}{n} + \frac{\alpha+\beta}{n+\alpha+\beta}\,\frac{\alpha}{\alpha+\beta} = \frac{x+\alpha}{n+\alpha+\beta}; \tag{14}$$

see [3, p. 299] and [5, p. 588]. Moreover, the weights associated with x/n and $\alpha/(\alpha+\beta)$ tend to 1 and 0, respectively, when n increases without bound. That is intuitively reasonable, since the influence of prior information about p should dissipate as we collect more and more data.

Example 2. The model used in this example, (1), has the same functional form as in (6) of Example 1, except that there are only two values for P and the weighing function (2) is different. The distribution of X given $Y = P$ in (1) is the conditional binomial probability mass function

$$f_{X\,|\,P}(x; n \mid p) = \binom{n}{x} p^x (1-p)^{n-x}, \tag{15}$$

for $x \in \{0, 1, 2, \ldots, n\}$ and $p \in \{p_1, p_2\}$, where $0 < p_1,\ p_2 < 1$, and $p_1 \neq p_2$. The distribution of parameter $Y = P$ in (2) is the discrete probability distribution

$$f_P(p; \theta) = \theta^{(p-p_2)/(p_1-p_2)}(1-\theta)^{(p_1-p)/(p_1-p_2)} = \begin{cases} \theta & \text{for } p = p_1, \\ 1-\theta & \text{for } p = p_2. \end{cases} \tag{16}$$

The joint distribution of X and P is the product of (15) and (16), i.e.

$$f_{X,P}(x, p; n, \theta) = \binom{n}{x} p^x (1-p)^{n-x} \theta^{(p-p_2)/(p_1-p_2)} (1-\theta)^{(p_1-p)/(p_1-p_2)}. \quad (17)$$

The marginal distribution of X is

$$f_X(x; n, \theta) = \sum_{p \in \{p_1, p_2\}} f_{X,P}(x, p; n, \theta) = \theta \binom{n}{x} p_1^x (1-p_1)^{n-x} + (1-\theta) \binom{n}{x} p_2^x (1-p_2)^{n-x}. \quad (18)$$

The conditional distribution of P given X is the quotient of (17) and (18):

$$f_{P \mid X}(p; n, \theta \mid x) = \frac{\binom{n}{x} p^x (1-p)^{n-x} \theta^{(p-p_2)/(p_1-p_2)} (1-\theta)^{(p_1-p)/(p_1-p_2)}}{\theta \binom{n}{x} p_1^x (1-p_1)^{n-x} + (1-\theta) \binom{n}{x} p_2^x (1-p_2)^{n-x}}, \quad (19)$$

for $p \in \{p_1, p_2\}$. The set of equations (15)–(19) is a realization of (1)–(5).

3. Applications

Each probability distribution in Sections 1 and 2 could serve as a probability model for a phenomenon or process. Equation (2) can be used in a variety of ways for different purposes. For example, further control of the shape and parameters in the model (1) can be achieved by extending (2) to a multistage or hierarchical modeling process; see [1, pp. 106–109, 180–195] and [5, pp. 606–613]. For instance, in Example 1 the binomial distribution's proportion p has a beta distribution in the pair (6) and (7), and additionally the beta distribution's parameter α could have a distribution, such as an exponential distribution.

Care must be taken when applying the sequence of formulas (1)–(5). The practitioner should be very sensitive to details about the random variables, data, and application, when choosing probability models. The applications below illustrate how (1)–(5) can be correctly used in some specific situations.

3.1. Mixture models

A mixture distribution is obtained in the general setting by integrating, or summing, the product (3) of (1) and (2) to obtain (4). One purpose is to generate distributions for variables from inputs that are meaningful to the experimenter. This is a tool for studying the genesis of a distribution of interest as the marginal distribution of (3). The mixture distributions in Examples 1 and 2 are (11) and (18), respectively. Other examples are a Poisson distribution with an Erlang distribution yields a negative binomial distribution and a geometric distribution with a beta distribution yields a Zipf distribution. For examples, see [5, pp. 190–195], [11, pp. 332–333], and [13, pp. 58–60, 83–84].

Equation (18) in Example 2 expresses a standard type of mixture of two distributions, in this case two binomial distributions; see [4, p. 181], [9, pp. 33–34, 181–185], and [10]. Often, normal distributions are the components of a mixture. The weighted sum in (18) is the distribution of the random variable X from two binomial processes that are sampled with probabilities θ and $1-\theta$. The weights might be taken from experimental evidence or initial educated guesses. In (16)–(18), the weights are fixed numbers. Sometimes, the weights are unknown values or random variables that must be estimated in a hierarchical model. The goal of the analysis might be to determine from observations the structure of a population that is believed to be composed of two populations, but it is unknown which population is the source of each measurement. These analyses were an important part of the early history of statistics and model building; see [12]. The populations could be sampled from different suppliers to a factory, various special interest groups in a survey, or different subspecies in a biological setting. The intent might be to discover how many subpopulations are present and in what proportions they occur, which are the weights. Mixtures of more than two populations are common. The literature on mixtures is large; see [9, pp. 40, 201] and [10].

3.2. Bayesian analyses

Equations (12) and (19) can be considered to be the distribution of P in Examples 1 and 2, after an experimental result $X = x$ has been obtained. The distribution of the parameter P is desired for a binomial probability model in which P initially has distribution (7) or (16). The focus is on the replacement of distributions (7) or (16) with (12) or (19), respectively. In Bayesian analysis, (2), (7), and (16) are called the *prior distributions* and (5), (12), and (19) are called the *posterior distributions*. In Example 1, the prior mean is (8) and the posterior mean is (13).

Equations (12)–(14) show how $\alpha + \beta$ from the beta distribution (7) plays a role like a sample size; see [2, pp. 146–147] and [5, p. 588]. This may help decide the values of α and β in the prior distribution (7). The larger the value of $\alpha + \beta$, the more impact the prior distribution (7) has on the posterior distribution (12), compared to the impact of the sample. Once the decision is made to use the beta prior distribution (7) with parameters α and β, the posterior distribution is the beta distribution (12), where the new parameters are obtained by adding the number of successes x to α and adding the number of failures $n - x$ to β; see [8, pp. 36–37]. The beta distribution is called the *conjugate prior distribution* for the binomial model, since the prior and posterior distributions have the same functional form; see [1, p. 130], [2, pp. 143–145, 155, 185–188], and [8, pp. 59–62].

The beta distribution (7) for P was chosen for many reasons, besides that it is the conjugate prior distribution. Its domain is the unit interval, which matches the domain of the binomial distribution's parameter p. It is very flexible since, by adjusting its two parameters, the distribution can be made to take many different shapes.

Selecting the uniform prior distribution, where $\alpha = 1$ and $\beta = 1$ in (7), expresses that there is no knowledge about P except that it can be any value between 0 and 1. The expected value (13) is

$$\frac{x+1}{n+2}, \tag{20}$$

which might be somewhat surprising, since it is different from the classical, nonBayesian, minimum variance unbiased estimator x/n for p (see [1, pp. 89–90] and [2, p. 155]), although the difference becomes negligible if n is large. Ross [11, pp. 141–143] presented alternative points of view on the probability modeling that produces (20).

4. Discussion

We have presented the case that the sequence of formulas (1)–(5) and the concepts that they contain are very useful. Facility with (1)–(5) and their ideas should be in the skill set of anyone using or studying probability or statistics.

In real situations, it often occurs that no standard mathematical form is known for a particular probability distribution. Rather, only observations are available. In some circumstances, when the functional form is available, the required integrations are intractable. In either event, although (1)–(5) still apply and the analysis can go forward, numerical methods may be required.

They have been extensively studied; see [5, pp. 286–296, 600–606], [6], [7], and [8, pp. 191–234].

References

[1] BERGER, J. O. (1993). *Statistical Decision Theory and Bayesian Analysis*, 2nd edn. Springer, New York.

[2] BOLSTAD, W. M. (2007). *Introduction to Bayesian Statistics,* 2nd edn. John Wiley, Hoboken, NJ.

[3] CASELLA, G. AND BERGER, R. L. (1990). *Statistical Inference*. Wadsworth & Brooks/Cole, Pacific Grove, CA.

[4] DEVORE, J. L. (2008). *Probability and Statistics for Engineering and the Sciences*, 7th edn. Thomson/Brooks/Cole, Belmont, CA.

[5] HOGG, R. V., MCKEAN, J. W. AND CRAIG, A. T. (2005). *Introduction to Mathematical Statistics*, 6th edn. Pearson, Upper Saddle River, NJ.

[6] JOHNSON, N. L., KOTZ, S. AND BALAKRISHNAN, N. (1994). *Continuous Univariate Distributions*, Vol. 1, 2nd edn. John Wiley, New York.

[7] JOHNSON, N. L., KOTZ, S. AND BALAKRISHNAN, N. (1995). *Continuous Univariate Distributions*, Vol. 2, 2nd edn. John Wiley, New York.

[8] KOCH, K.-R. (2007). *Introduction to Bayesian Statistics*, 2nd edn. Springer, Berlin.

[9] LAWLESS, J. F. (2003). *Statistical Models and Methods for Lifetime Data*, 2nd edn. John Wiley, Hoboken, NJ.

[10] MCLACHLAN, G. AND PEEL, D. (2000). *Finite Mixture Models*. John Wiley, New York.

[11] ROSS, S. M. (2014). *Introduction to Probability Models*, 11th edn. Elsevier, Amsterdam.

[12] STIGLER, S. M. (1986). *The History of Statistics: The Measurement of Uncertainty Before 1900*. Belknap, Cambridge, MA.

[13] TAYLOR, H. M. AND KARLIN, S. (1998). *An Introduction to Stochastic Modeling*, 3rd edn. Academic Press, San Diego, CA.