Wiley StatsRef: Statistics Reference Online

Mixture of Distributions

By *Harry H. Panjer*

Keywords: *mixed distributions, mixed* Poisson*, negative binomial, binomial-beta, negative hypergeometric,* Polya–Eggenberger*, generalized* Waring*, mixed exponential, frailty models*

Abstract: This article gives an overview of some probability distributions that are obtained as mixtures of known distributions.

In this article, we examine **mixing distributions** by treating one or more parameters as being 'random' in some sense. This idea is discussed in connection with the **mixtures of the Poisson distribution**.

We assume that the parameter of a probability distribution is itself distributed over the population under consideration (the 'collective') and that the sampling scheme that generates our data has two stages. First, a value of the parameter is selected from the distribution of the parameter. Then, given the selected parameter value, an observation is generated from the population using that parameter value.

In **automobile insurance**, for example, classification schemes attempt to put individuals into (relatively) homogeneous groups for the purpose of pricing. Variables used to develop the classification scheme might include age, experience, a history of violations, accident history, and other variables. Since there will always be some residual variation in accident risk within each class, mixed distributions provide a framework for modeling this heterogeneity.

For **claim size distributions**, there may be uncertainty associated with future claims inflation, and scale mixtures often provide a convenient mechanism for dealing with this uncertainty.

Furthermore, for both **discrete** and **continuous** distributions, mixing also provides an approach for the construction of alternative models that may well provide an improved fit to a given set of data.

Let $M(t|\theta) = \int_0^\infty \mathrm{e}^{tx} f(x|\theta)\mathrm{d}x$ denote the **moment generating function** (mgf) of the probability distribution, if the risk parameter is known to be θ. The parameter, θ, might be the **Poisson** mean, for example, in which case the measurement of risk is the expected number of events in a fixed time period.

Let $U(\theta) = \Pr(\Theta \leq \theta)$ be the cumulative distribution function (cdf) of Θ, where Θ is the risk parameter, which is viewed as a random variable. Then $U(\theta)$ represents the probability that, when a value of Θ is selected (e.g. a driver is included in the automobile example), the value of the risk parameter does not

University of Waterloo, Department of Statistics and Actuarial Science, Waterloo, Ontario, Canada N2L 3G1
(E-mail: hpanjer@uwaterloo.ca)

This article was originally published online in 2006 in Encyclopedia of Actuarial Science, © John Wiley & Sons, Ltd and republished in Wiley StatsRef: Statistics Reference Online, 2014.

exceed θ. Then,

$$M(t) = \int M(t|\theta)\mathrm{d}U(\theta) \tag{1}$$

is the unconditional mgf of the probability distribution. The corresponding unconditional probability distribution is denoted by

$$f(x) = \int f(x|\theta)\mathrm{d}U(\theta). \tag{2}$$

The mixing distribution denoted by $U(\theta)$ may be of the discrete or continuous type or even a combination of discrete and continuous types. Discrete mixtures are mixtures of distributions when the mixing function is of the discrete type. Similarly for continuous mixtures.

It should be noted that the mixing distribution is normally unobservable in practice, since the data are drawn only from the mixed distribution.

Example 1. The **zero-modified distributions** may be created by using two-point mixtures since

$$M(t) = p1 + (1 - p)M(t|\theta). \tag{3}$$

This is a (discrete) two-point mixture of a **degenerate distribution** (i.e. all probability at zero), and the distribution with mgf $M(t|\theta)$.

Example 2. Suppose drivers can be classified as 'good drivers' and 'bad drivers', each group with its own **Poisson distribution**. This model and its application to the data set are from Tröbliger [4].

From (2) the unconditional probabilities are

$$f(x) = p\frac{\mathrm{e}^{-\lambda_1}\lambda_1^x}{x!} + (1-p)\frac{\mathrm{e}^{-\lambda_2}\lambda_2^x}{x!}, \quad x = 0, 1, 2, \ldots. \tag{4}$$

Maximum likelihood estimates of the parameters were calculated by Tröbliger [4] to be $\hat{p} = 0.94$, $\hat{\lambda}_1 = 0.11$, and $\hat{\lambda}_2 = 0.70$. This means that about 6% of drivers were 'bad' with a risk of $\lambda_1 = 0.70$ expected accidents per year and 94% were 'good' with a risk of $\lambda_2 = 0.11$ expected accidents per year.

Mixed Poisson distributions are discussed in detail in another article of the same name in this encyclopedia. Many of these involve continuous mixing distributions. The most well-known mixed **Poisson distribution** includes the **negative binomial** (Poisson mixed with the **gamma distribution**) and the Poisson-inverse Gaussian (Poisson mixed with the **inverse Gaussian distribution**).

Many other mixed models can be constructed beginning with a simple distribution.

Example 3. **Binomial** mixed with a **beta distribution**. This distribution is called *binomial-beta*, negative hypergeometric, or Polya–Eggenberger.

The beta distribution has probability density function

$$u(q) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\, q^{a-1}(1-q)^{b-1}, \quad a > 0, b > 0, 0 < q < 1. \tag{5}$$

Then the mixed distribution has probabilities

$$
\begin{aligned}
f(x) &= \int_0^1 \binom{m}{x} q^x (1-q)^{m-x} \\
&\quad \times \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} q^{a-1}(1-q)^{b-1}\,\mathrm{d}q \\
&= \frac{\Gamma(a+b)\Gamma(m+1)\Gamma(a+x)\Gamma(b+m-x)}{\Gamma(a)\Gamma(b)\Gamma(x+1)\ \ \Gamma(m-x+1)\Gamma(a+b+m)} \\
&= \frac{\binom{-a}{x}\binom{-b}{m-x}}{\binom{-a-b}{m}}, x = 0, 1, 2, \ldots.
\end{aligned}
\tag{6}
$$

Example 4. Negative binomial distribution mixed on the parameter $p = (1 + \beta)^{-1}$ with a **beta distribution**. The mixed distribution is called the *generalized Waring*.

Arguing as in Example 3 , we have

$$
\begin{aligned}
f(x) &= \frac{\Gamma(r+x)}{\Gamma(r)\Gamma(x+1)} \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \\
&\quad \times \int_0^1 p^{a+r-1}(1-p)^{b+x-1}\,\mathrm{d}p \\
&= \frac{\Gamma(r+x)}{\Gamma(r)\Gamma(x+1)} \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \frac{\Gamma(a+r)\Gamma(b+x)}{\Gamma(a+r+b+x)}, \\
&\qquad x = 0, 1, 2, \ldots.
\end{aligned}
\tag{7}
$$

When $b = 1$, this distribution is called the *Waring distribution*. When $r = b = 1$, it is termed the ***Yule distribution***.

Examples 3 and 4 are mixtures. However, the mixing distributions are not **infinitely divisible**, because the beta distribution has finite support. Hence, we cannot write the distributions as compound Poisson distributions to take advantage of the compound Poisson structure.

The following class of mixed continuous distributions is of central importance in many areas of actuarial science.

Example 5. Mixture of exponentials

The mixed distribution has probability density function

$$
f(x) = \int_0^\infty \theta\, \mathrm{e}^{-\theta x}\,\mathrm{d}U(\theta), x > 0. \tag{8}
$$

Let $M_\Theta(t) = \int_0^\infty \mathrm{e}^{t\theta}\,\mathrm{d}U(\theta)$ be the mgf associated with the random parameter θ, and one has $f(x) = M_\Theta{}'(-x)$. For example, if θ has the **gamma** probability density function

$$
u(\theta) = \frac{\lambda(\lambda\theta)^{\alpha-1}\,\mathrm{e}^{-\lambda\theta}}{\Gamma(\alpha)}, \theta > 0, \tag{9}
$$

then $M_\Theta(t) = [\Theta/(\Theta - t)]^\alpha$, and thus

$$f(x) = \alpha\lambda^\alpha(\lambda + x)^{-\alpha-1}, \ x > 0, \tag{10}$$

a **Pareto distribution**. All exponential mixtures, including the Pareto, have a decreasing failure rate. Moreover, exponential mixtures form the basis upon which the so-called **frailty** models are formulated.

Excellent references for in-depth treatment of mixed distributions are [1–3].

Related Articles

Capital Allocation for P&C Insurers: A Survey of Methods
Compound Process
Failure Rate
Hidden Markov Models
Lundberg Inequality for Ruin Probability
Markov Chain Monte Carlo Methods
Nonexpected Utility Theory
Phase-Type Distributions
Phase Method
Replacement Value
Sundt's Classes of Distributions
Under- and Overdispersion
Value-at-Risk

References

[1] Johnson, N., Kotz, S. & Balakrishnan, N. (**1994**). *Continuous Univariate Distributions*, Vol. *1*, 2nd Edition, Wiley, New York.
[2] Johnson, N., Kotz, S. & Balakrishnan, N. (**1994**). *Continuous Univariate Distributions*, Vol. *2*, 2nd Edition, Wiley, New York.
[3] Johnson, N., Kotz, S. & Kemp, A. (**1992**). *Univariate Discrete Distributions*, 2nd Edition, Wiley, New York.
[4] Tröbliger, A. (**1961**). Mathematische Untersuchungen zur Beitragsrückgewähr in der Kraftfahrversicherung, *Blätter der Deutsche Gesellschaft für Versicherungsmathematik* *5*, 327–348.