Section 3.7. Mixture Distributions

Note. At the end of Section 3.4. The Normal Distribution we saw an example of a linear combination of normal distributions (in the setting of cumulative distribution functions). We follow up on this idea in this section by considering weighted sums of probability density functions.

Definition. Let $f_1(x), f_2(x), \ldots, f_k(x)$ be probability density functions with supports $\mathcal{S}_1, \mathcal{S}_2, \ldots, \mathcal{S}_k$, with means $\mu_1, \mu_2, \ldots, \mu_k$, and standard deviations $\sigma_1^2, \sigma_2^2, \ldots, \sigma_k^2$, respectively. Let positive probabilities $p_1, p_2, \ldots, p_k$ satisfy $p_1 + p_2 + \cdots + p_k = 1$. Let $\mathcal{S} = \cup_{i=1}^k \mathcal{S}_i$. The function

$$f(x) = p_1 f_1(x) + p_2 f_2(x) + \cdots + p_k f_k(x) = \sum_{i=1}^k p_i f_i(x) \text{ for } x \in \mathcal{S}$$

is a *mixture distribution* with *mixing probabilities* $p_1, p_2, \ldots, p_k$.

Note. By additivity of the integral, we clearly have $\int_{-\infty}^{\infty} f(x)\,dx = 1$. Also with $F_i(x)$ as the cumulative density function corresponding to $f_i(x)$, then the cumulative distribution function of f is $F(x) = \sum_{i=1}^k p_i F_i(x)$ for $x \in \mathcal{S}$. The mean of X is

$$E(X) = \sum_{i=1}^k p_i \int_{-\infty}^{\infty} x f_i(x)\,dx = \sum_{i=1}^k p_i \mu_i = \overline{\mu},$$

and the variance is

$$\text{var}(X) = \sum_{i=1}^k p_i \int_{-\infty}^{\infty} (x - \overline{\mu})^2 f_i(x)\,dx$$

$$\begin{aligned}
&= \sum_{i=1}^{k} p_i \int_{-\infty}^{\infty} ((x-\mu_i)+(\mu_i-\overline{\mu}))^2 f_i(x)\,dx \\
&= \sum_{i=1}^{k} p_i \int_{-\infty}^{\infty} (x-\mu_i)^2 f_i(x)\,dx + \sum_{i=1}^{k} 2p_i(\mu_i-\overline{\mu}) \int_{-\infty}^{\infty} (x-\mu_i) f_i(x)\,dx \\
&\quad + \sum_{i=1}^{k} p_i(\mu_i=\overline{\mu})^2 \int_{-\infty}^{\infty} f_i(x)\,dx \\
&= \sum_{i=1}^{k} p_i \int_{-\infty}^{\infty} (x-\mu_i)^2 f_i(x)\,dx + \sum_{i=1}^{k} p_i(\mu_i=\overline{\mu})^2(1) \\
&\quad \text{since } \int_{-\infty}^{\infty} (x-\mu_i) f_i(x)\,dx = \int_{-\infty}^{\infty} x f_i(x)\,dx - \mu_i \int_{-\infty}^{\infty} f_i(x)\,dx \\
&\quad = \mu_i - \mu_i = 0 \\
&= \sum_{i=1}^{k} p_i\sigma_i^2 + \sum_{i=1}^{k} p_i(\mu_i-\overline{\mu})^2.
\end{aligned}$$

Definition. Random variable X has a *loggamma* probability density function with parameters $\alpha > 0$ and $\beta > 0$ if it has probability density function

$$f(x) = \begin{cases} \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{-(1+\beta)/\beta} (\log x)^{\alpha-1} & \text{for } x > 1 \\ 0 & \text{elsewhere.} \end{cases}$$

We denote this distribution as $\log\Gamma(\alpha,\beta)$.

Note. Here, “log” denotes the natural logarithm function. The mean of $\log\Gamma(\alpha,\beta)$ is $(1-\beta)^{-\alpha}$ when $\beta < 1$, and variance $(1-2\beta)^{-\alpha} - (1-\beta)^{-2\alpha}$ when $\beta < 1/2$ (see Vose Software’s loggamma webpage).

Note. We can also consider a mixture of an infinite number of distributions. This is accomplished by integrating instead of summing (and so involves an uncountable infinity of distributions; we get these distributions by varying a continuous parameter in a class of distributions).

Example 3.7.2. Let X_θ be a Poisson random variable with parameter θ (so the probability mass function of X_θ is $\theta^x e^{-\theta}/x!$ for $x \in \{0,1,2,\ldots\}$). As opposed to weighing by a probability p_i, we introduce a weight function. We use a gamma distribution with parameters α and β (so, as a function of θ, we use $\dfrac{1}{\Gamma(\alpha)\beta^\alpha}\theta^{\alpha-1}e^{-\theta/\beta}$ for $0 < \theta < \infty$). We then get the probability mass function

$$\begin{aligned} p(x) &= \int_0^\infty \left(\frac{1}{\Gamma(\alpha)\beta^\alpha}\theta^{\alpha-1}e{-}\theta/\beta\right)\left(\frac{\theta^x e^{-\theta}}{x!}\right) d\theta \\ &= \frac{1}{\Gamma(\alpha)\beta^\alpha x!}\int_0^\infty \theta^{\alpha+x-1}e^{-\theta(1+\beta)/\beta}\, d\theta \\ &= \frac{\Gamma(\alpha+x)\beta^{\alpha+x}}{\Gamma(\alpha)\beta^\alpha x!(1+\beta)^{\alpha+x}} = \frac{\Gamma(\alpha+x)\beta^x}{\Gamma(\alpha)x!(1+\beta)^{\alpha+x}} \end{aligned}$$

since $\displaystyle\int_0^\infty \frac{1}{\Gamma(\alpha+x)}\left(\frac{\beta}{1+\beta}\right)^{\alpha+x}\theta^{\alpha+x-1}e^{0\theta(1+\beta)/\beta}\, d\theta = 1$ because the integrand is the probability density function of a gamma distribution with parameters $\alpha + x > 0$ and $\dfrac{\beta}{1+\beta} > 0$.

In the event that we take $\alpha = r \in \mathbb{N}$ and $\beta = (1-p)/p$ where $0 < p < 1$ then probability density function $p(x)$ becomes

$$p(x) = \frac{\Gamma(r+x)((1-p)/p)^x}{\Gamma(r)x!(1/p)^{r+x}} = \frac{(r+x-1)!}{(r-1)!}\frac{p^r 1-p)^x}{x!} \text{ for } x \in \{0,1,2,\ldots\}.$$

This is related to Bernoulli trials. It gives the probability of r “successes” in the performance of $r + x$ trials; so variable x represents the “number of excess trials needed to obtain r successes in a sequence of independent [Bernoulli] trials." (See Hogg, McKean, and Craig page 220.) This one form of the "negative binomial distribution."

Note. In the previous example, we have every right to express concern over whether or not $p(x)$ actually sums to 1. We resolve this by considering $X = X_\theta$ as a conditional distribution given θ, so that we have $X_\theta(x) = f(x \mid \theta)$. Then the weighting function is treated as a probability density function for θ, say $g(\theta)$. The joint probability density function of θ and x is $g(\theta)f(x \mid \theta)$. The mixture (or "compound") probability density function can then be thought of as the marginal (or "unconditional") probability density function of X, $h(x) = \int_\theta g(\theta)f(x \mid \theta)\, d\theta$. In the event that θ has a discrete distribution, the integral is replaced with a sum (or series).

Example 3.7.4. Suppose X has a conditional gamma probability density function with parameters $\alpha = k$ and $\beta = \theta^{-1}$, so that

$$X = X_\theta(x) = \frac{1}{\Gamma(k)(\theta^{-1})^k} x^{k-1} e^{-x/(\theta^{-1})} = \frac{\theta^k x^{k-1} e^{-\theta x}}{\Gamma(k)}.$$

Take the weighting function of θ as a gamma probability density function with parameters α and β, so it is $\dfrac{1}{\Gamma(\alpha)\beta^\alpha}\theta^{\alpha-1}e^{-\theta/\beta}$. Then the unconditional (compound) probability density function of X is

$$\begin{aligned} h(x) &= \int_0^\infty \left(\frac{\theta^{\alpha-1}e^{-\theta/\beta}}{\Gamma(\alpha)\beta^\alpha}\right)\left(\frac{\theta^k x^{k-1} e^{-\theta x}}{\Gamma(k)}\right) d\theta \\ &= \int_0^\infty \frac{x^{k-1}\theta^{\alpha+k-1}}{\beta^\alpha \Gamma(\alpha)\Gamma(k)} e^{-\theta(1+\beta x)/\beta}\, d\theta \end{aligned}$$

$$
\begin{aligned}
&= \frac{x^{k-1}\Gamma(\alpha+k)(\beta/(1+\beta x))^{\alpha+k}}{\beta^\alpha\Gamma(\alpha)\Gamma(k)} \\
&\quad \times \int_0^\infty \frac{1}{\Gamma(\alpha+k)(\beta/(1+\beta x))^{\alpha+k}}\theta^{(\alpha+k)-1}e^{-\theta(1+\beta x)/\beta}\,d\theta \\
&= \frac{\Gamma(\alpha+k)\beta^k x^{k-1}}{\Gamma(\alpha)\Gamma(k)(1+\beta x)^{\alpha+k}}(1) \text{ since the integrand above is the pdf} \\
&\quad \text{for a } \Gamma \text{ distribution with parameters } \alpha+k \text{ and } \beta/(1+\beta x) \\
&= \frac{\Gamma(\alpha+k)\beta^k x^{k-1}}{\Gamma(\alpha)\Gamma(k)(1+\beta x)^{\alpha+k}} \text{ for } 0 < x < \infty.
\end{aligned}
$$

This probability density function is the *generalized Pareto distribution*. With $k = 1$ we have

$$h(x) = \alpha\beta(1+\beta x)^{-(\alpha+1)} \text{ for } 0 < x < \infty,$$

known as the *Pareto distribution*. The Pareto probability density functions have "thicker tails" than the gamma distributions which act as the conditional distribution (Hogg, McKean, and Craig claim; see page 222).

Note. The cumulative distribution function of the Pareto probability density function is

$$
\begin{aligned}
H(x) &= \int_0^x \alpha\beta(1+\beta t)^{-(\alpha+1)}\,dt = \frac{\alpha\beta}{(-\alpha)(\beta)}(1+\beta t)^{-\alpha}\Big|_0^x \\
&= -(1+\beta x)^{-\alpha} + 1 = 1 - (1+\beta x)^{-\alpha} \text{ for } 0 \le x < \infty.
\end{aligned}
$$

If we let $X = Y^r$ where $r > 0$ in the Pareto probability density function then the cumulative distribution function of Y is

$$G(y) = P(Y \le y) = P(X^{1/r} \le y) = P(X \le y^r)$$

$$= H(y^r) = 1 - (1+\beta y^r)^{-\alpha} \text{ for } 0 < y < \infty$$

and so the probability density function of y is, by Note 1.7.A,

$$g(y) = G'(y) = -(-\alpha)(\beta r y^{r-1})(1+\beta y^r)^{-\alpha-1} = \frac{\alpha\beta r y^{r-1}}{(1+\beta y^r)^{\alpha+1}} \text{ for } 0 < y < \infty.$$

The associated distribution with this probability density function is the *transformed Pareto distribution* or the *Burr distribution*.

Revised: 7/22/2021