Statistical Machine Learning Distributions

Lecture Notes 4: Multinomial Distribution

Professor: Zhihua Zhang

2.3.1 More About Mixture Distribution

Definition 2.1. *In probability theory and statistics, the moment-generating function of a random variable X is*

$$M_X(t) = \mathbb{E}[e^{tX}] = \int e^{tx} f_X(x)dx$$

One property about moment-generating function is that we can get $\mathbb{E}[X^k]$ from $M_X^{(k)}(0)$, as we can see $M_X^{(k)}(t) = \int x^k e^{tx} f_X(x)dx$, where we assume we can put the derivation inside. So $M_X^{(k)}(0) = \mathbb{E}[X^k]$.

Definition 2.2. *A function $f : (0, \infty) \to \mathbb{R}$ is completely monotone function if and only if f is of class C^∞(infinitely derivable), and $(-1)^n f^{(n)}(\lambda) \geq 0$ for all $n \in N \cup \{0\}$, and $\lambda > 0$.*

Theorem 2.1. *(Bernstein) Let $g : (0, \infty) \to \mathbb{R}$ be a completely monotone function. Then it is the Laplace transform of unique measure μ on $[0, \infty]$, i.e. for all $\lambda > 0$,*

$$g(\lambda) = \mathcal{L}(\mu; \lambda) = \int_{[0,\infty)} e^{-\lambda t} \mu(dt)$$

. Conversely, whenever $\mathcal{L}(\mu; \lambda) < \infty$ for every $\lambda > 0$, $\lambda \mapsto \mathcal{L}(\mu; \lambda)$ is a completely monotone function.

Proof. Assume $g(0+) = 1$ and $g(+\infty) = 0$. By Taylor's formula

$$\begin{aligned} f(\lambda) &= \sum_{k=0}^{n-1} \frac{f^{(k)}(a)}{k!}(\lambda - a)^k + \int_a^\lambda \frac{f^{(n)}(s)}{(n-1)!}(\lambda - s)^{n-1} ds \\ &= \sum_{k=0}^{n-1} \frac{(-1)^k f^{(k)}(a)}{k!}(a - \lambda)^k + \int_\lambda^a \frac{(-1)^n f^{(n)}(s)}{(n-1)!}(s - \lambda)^{n-1} ds \end{aligned} \tag{1}$$

where $a > 0$ and $n \in \mathbb{N}$. Let $a \to \infty$, then

$$\begin{aligned} \lim_{a \to \infty} \int_\lambda^a \frac{(-1)^n f^{(n)}(s)}{(n-1)!}(s - \lambda)^{n-1} ds &= \int_\lambda^\infty \frac{(-1)^n f^{(n)}(s)}{(n-1)!}(s - \lambda)^{n-1} ds \\ &\leq f(\lambda). \end{aligned}$$

So the sum in (1) converges for every $n \in \mathbb{N}$ as $a \to \infty$. Let

$$\rho_n(\lambda) = \lim_{a \to \infty} \frac{(-1)^n f^{(n)}(a)}{n!}(a - k)^n$$

. This limit doesn't depend on $\lambda > 0$. Indeed, for $k > 0$,

$$\begin{aligned}
\rho_n(k) &= \lim_{a\to\infty} \frac{(-1)^n f^{(n)}(a)}{n!}(a-k)^n \\
&= \lim_{a\to\infty} \frac{(-1)^n f^{(n)}(a)}{n!}(a-\lambda)^n \frac{(a-k)^n}{(a-\lambda)^n} \\
&= \rho_n(\lambda).
\end{aligned}$$

So we can get

$$f(\lambda) = \sum_{k=0}^{n-1} \rho_k(\lambda) + \int_\lambda^\infty \frac{(-1)^n f^{(n)}(s)}{(n-1)!}(s-\lambda)^{n-1} ds$$

Let $\lambda \to \infty$, since $f(+\infty) = 0$, so $\rho_k(\lambda) = 0$. Then we can get

$$f(\lambda) = \int_\lambda^\infty \frac{(-1)^n f^{(n)}(s)}{(n-1)!}(s-\lambda)^{n-1} ds \tag{2}$$

. And since $f(0+) = 1$, we can get:

$$1 = \lim_{\lambda\to 0+} f(\lambda) = \int_0^\infty \frac{(-1)^n f^{(n)}(s)}{(n-1)!} s^{n-1} ds$$

And (2) can also be written as:

$$f(\lambda) = \int_0^\infty (1-\frac{\lambda}{s})_+^{n-1} \frac{(-1)^n f^{(n)}(s)}{(n-1)!} s^{n-1} ds.$$

Let $t = \frac{n}{s}$, then

$$f(\lambda) = \int_0^\infty (1-\frac{\lambda t}{n})_+^{n-1} \frac{(-1)^n}{n!} f^{(n)}(\frac{n}{t})(\frac{n}{t})^{n+1} dt$$

. Since $\lim_{n\to\infty}(1-\frac{\lambda t}{n})_+^{n-1} = e^{-\lambda t}$. So

$$f(\lambda) = \int_0^\infty e^{-\lambda t} \frac{(-1)^n}{n!} f^{(n)}(\frac{n}{t})(\frac{n}{t})^{n+1} dt.$$

For the converse, let $f(\lambda) = \mathcal{L}(\mu;\lambda) = \int_0^\infty e^{-\lambda t}\mu(dt)$. So

$$(-1)^n f^{(n)}(\lambda) = \int_0^\infty t^n e^{-\lambda t}\mu(dt) \geq 0$$

$\square$