Chapter 2 Probability Distributions

Yue Yu

2.1 Binary Variables

(2.3)

$$
\begin{aligned}
\mathbb{E}[\mathbf{x}] &= \sum_x x p(x) \\
&= \sum_x x \mu^x (1-\mu)^{1-x} \\
&= 1 \cdot \mu^1 (1-\mu)^0 + 0 \cdot \mu^0 (1-\mu)^1 \\
&= \mu.
\end{aligned}
$$

(2.4)

$$
\begin{aligned}
\mathrm{var}[\mathbf{x}] &= \mathbb{E}[x^2] - \mathbb{E}[x]^2 \\
&= \sum_x x^2 p(x) - \mu^2 \\
&= \mu(1-\mu).
\end{aligned}
$$

(2.7)

$$
\begin{aligned}
\mu_{\mathrm{ML}} &= \arg_\mu \left(\frac{\partial}{\partial \mu} \ln p(\mathcal{D}|\mu) = 0 \right) \\
&= \arg_\mu \left(\frac{\partial}{\partial \mu} \sum_{n=1}^{N} \{ x_n \ln \mu + (1-x_n) \ln(1-\mu) \} = 0 \right) \\
&= \arg_\mu \left(\sum_{n=1}^{N} \frac{x_n}{\mu} - \sum_{n=1}^{N} \frac{1-x_n}{1-\mu} = 0 \right) \\
&= \frac{1}{N} \sum_{n=1}^{N} x_n.
\end{aligned}
$$

(2.11) and (2.12)

We prove these two equations through problem 2.3 and 2.4. Firstly, notice that

$$\begin{aligned}\binom{N}{m}+\binom{N}{m-1} &= \frac{N!}{(N-m)!m!}+\frac{N!}{(N-m+1)!(m-1)!} \\ &= \frac{N!(N-m+1)}{(N-m+1)!m!}+\frac{N!m}{(N-m+1)!m!} \\ &= \frac{(N+1)!}{(N+1-m)!m!} \\ &= \binom{N+1}{m}. \qquad (*)\end{aligned}$$

Now, we prove by induction the *binomial theorem* that is given by

$$(1+x)^N=\sum_{m=0}^{N}\binom{N}{m}x^m.$$

When $N=0$, we have

$$\sum_{m=0}^{0}\binom{0}{m}x^m=1=(1+x)^0.$$

If the equation is correct for any integer $N>0$, then for $N+1$, we have

$$\begin{aligned}(1+x)^{N+1} &= (1+x)^N(1+x) \\ &= \sum_{m=0}^{N}\binom{N}{m}x^m(1+x) \\ &= \sum_{m=0}^{N}\binom{N}{m}x^m+\sum_{m=0}^{N}\binom{N}{m}x^{m+1} \\ &= \binom{N}{0}x^0+\sum_{m=1}^{N}\binom{N}{m}x^m+\sum_{m=0}^{N-1}\binom{N}{m}x^{m+1}+\binom{N}{N}x^{N+1} \\ &= \binom{N+1}{0}x^0+\left[\sum_{m=1}^{N}\binom{N}{m}x^m+\sum_{m=1}^{N}\binom{N}{m-1}x^m\right]+\binom{N+1}{N+1}x^{N+1} \\ &= \binom{N+1}{0}x^0+\sum_{m=1}^{N}\binom{N+1}{m}x^m+\binom{N+1}{N+1}x^{N+1} \\ &= \sum_{m=0}^{N+1}\binom{N+1}{m}x^m,\end{aligned}$$

where the fifth step used (*). Hence, the binomial theorem holds.

Next, we prove the binomial distribution is normalized. Specifically,

$$\begin{aligned}\sum_{m=0}^{N}\mathrm{Bin}(m|N,\mu) &= \sum_{m=0}^{N}\binom{N}{m}\mu^m(1-\mu)^{N-m} \\ &= (1-\mu)^N\sum_{m=0}^{N}\binom{N}{m}\left(\frac{\mu}{1-\mu}\right)^m \\ &= (1-\mu)^N\left(1+\frac{\mu}{1-\mu}\right)^N \\ &= 1, \qquad (*)\end{aligned}$$

where the second last step used the binomial theorem that we just proved.

Differentiating both sides of (*) with respect to μ, we have

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}\mu}\sum_{m=0}^{N}\mathrm{Bin}(m|N,\mu) &= \sum_{m=0}^{N}\binom{N}{m}\left(m\mu^{m-1}(1-\mu)^{N-m}-(N-m)\mu^{m}(1-\mu)^{N-m-1}\right)\\
&= \sum_{m=0}^{N}\binom{N}{m}\left(\frac{m}{\mu}\mu^{m}(1-\mu)^{N-m}-\frac{N-m}{1-\mu}\mu^{m}(1-\mu)^{N-m}\right)\\
&= \frac{1}{\mu(1-\mu)}\sum_{m=0}^{N}(m-N\mu)\binom{N}{m}\mu^{m}(1-\mu)^{N-m}\\
&= \frac{1}{\mu(1-\mu)}\sum_{m=0}^{N}(m-N\mu)\mathrm{Bin}(m|N,\mu)\\
&= 0.
\end{aligned}$$

Rearranging the equation, we obtain

$$\begin{aligned}
\mathbb{E}[m] &= \sum_{m=0}^{N} m\mathrm{Bin}(m|N,\mu)\\
&= N\mu\sum_{m=0}^{N}\mathrm{Bin}(m|N,\mu)\\
&= N\mu,
\end{aligned}$$

where we used the fact we just proved that the binomial distribution is normalized.

To compute the variance, we further differentiate both sides of the above equation with respect to μ,

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}\mu}\mathbb{E}[m] &= \sum_{m=0}^{N} m\binom{N}{m}\left(m\mu^{m-1}(1-\mu)^{N-m}-(N-m)\mu^{m}(1-\mu)^{N-m}\right)\\
&= \frac{1}{\mu(1-\mu)}\sum_{m=0}^{N}(m^2-mN\mu)\binom{N}{m}u^{m}(1-\mu)^{N-m}\\
&= \frac{1}{\mu(1-\mu)}\sum_{m=0}^{N}(m^2-mN\mu)\mathrm{Bin}(m|N,\mu)\\
&= \frac{1}{\mu(1-\mu)}\left\{\sum_{m=0}^{N}m^2\mathrm{Bin}(m|N,\mu)-N\mu\sum_{m=0}^{N}m\mathrm{Bin}(m|N,\mu)\right\}\\
&= \frac{1}{\mu(1-\mu)}(\mathbb{E}[m^2]-\mathbb{E}[m]^2)\\
&= \frac{1}{\mu(1-\mu)}\mathrm{var}[m]\\
&= N.
\end{aligned}$$

Therefore,

$$\mathrm{var}[m] = N\mu(1-\mu).$$