

## Chapter 2 Probability Distributions

Yue Yu

### 2.1 Binary Variables

(2.3)

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

(2.4)

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

(2.7)

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

**(2.11) and (2.12)**

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

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

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,

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

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

Differentiating both sides of (\*) with respect to  $\mu$ , we have

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

Rearranging the equation, we obtain

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

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  $\mu$ ,

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

Therefore,

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

## (2.14)

From the definition of the gamma function

$$\Gamma(x) = \int_0^\infty u^{x-1} e^{-u} du,$$

we have

$$\begin{aligned}\Gamma(a)\Gamma(b) &= \int_0^\infty \exp(-x)x^{a-1} dx \int_0^\infty \exp(-y)y^{b-1} dy \\ &= \int_0^\infty \int_0^\infty \exp(-(x+y))x^{a-1}y^{b-1} dy dx.\end{aligned}$$

Substituting  $t = x + y$ , we have

$$\begin{aligned}\Gamma(a)\Gamma(b) &= \int_0^\infty \int_x^\infty \exp(-t)x^{a-1}(t-x)^{b-1} \left| \frac{dy}{dt} \right| dt dx \\ &= \int_0^\infty \int_x^\infty \exp(-t)x^{a-1}(t-x)^{b-1} dt dx \\ &= \int_0^\infty \int_0^t \exp(-t)x^{a-1}(t-x)^{b-1} dx dt.\end{aligned}$$

We further substitute  $x = t\mu$ , which gives

$$\begin{aligned}\Gamma(a)\Gamma(b) &= \int_0^\infty \int_0^1 \exp(-t)(t\mu)^{a-1}(t-t\mu)^{b-1} \left| \frac{dx}{d\mu} \right| d\mu dt \\ &= \int_0^\infty \int_0^1 \exp(-t)(t\mu)^{a-1}(t-t\mu)^{b-1} t d\mu dt \\ &= \int_0^\infty \exp(-t)t^{a+b-1} dt \int_0^1 \mu^{a-1}(1-\mu)^{b-1} d\mu \\ &= \Gamma(a+b) \int_0^1 \mu^{a-1}(1-\mu)^{b-1} d\mu.\end{aligned}$$

Therefore,

$$\begin{aligned}\int_0^1 \text{Beta}(\mu|a, b) d\mu &= \int_0^1 \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \mu^{a-1}(1-\mu)^{b-1} d\mu \\ &= 1.\end{aligned}$$

**(2.15)**

$$\begin{aligned}\mathbb{E}[\mu] &= \int_0^1 \mu \text{Beta}(\mu|a, b) d\mu \\ &= \int_0^1 \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \mu^a(1-\mu)^{b-1} d\mu \\ &= \frac{\Gamma(a+b)\Gamma(a+1)}{\Gamma(a+b+1)\Gamma(a)} \int_0^1 \frac{\Gamma(a+b+1)}{\Gamma(a+1)\Gamma(b)} \mu^a(1-\mu)^{b-1} d\mu \\ &= \frac{\Gamma(a+b)\Gamma(a+1)}{\Gamma(a+b+1)\Gamma(a)} \\ &= \frac{a\Gamma(a+b)\Gamma(a)}{(a+b)\Gamma(a+b)\Gamma(a)} \\ &= \frac{a}{a+b},\end{aligned}$$

where the third step used the fact that the gamma distribution is normalized, and the second last step used the property  $\Gamma(x+1) = x\Gamma(x)$ .

(2.16)

$$\begin{aligned}\text{var}[\mu] &= \mathbb{E}[\mu^2] - \mathbb{E}[\mu]^2 \\&= \int_0^1 \mu^2 \text{Beta}(\mu|a, b) \, d\mu - \left(\frac{a}{a+b}\right)^2 \\&= \int_0^1 \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \mu^{a+1} (1-\mu)^{b-1} \, d\mu - \left(\frac{a}{a+b}\right)^2 \\&= \frac{\Gamma(a+b)\Gamma(a+2)}{\Gamma(a+b+2)\Gamma(a)} \int_0^1 \frac{\Gamma(a+b+2)}{\Gamma(a+2)\Gamma(b)} \mu^{a+1} (1-\mu)^{b-1} \, d\mu - \left(\frac{a}{a+b}\right)^2 \\&= \frac{\Gamma(a+b)\Gamma(a+2)}{\Gamma(a+b+2)\Gamma(a)} - \left(\frac{a}{a+b}\right)^2 \\&= \frac{a(a+1)\Gamma(a+b)\Gamma(a)}{(a+b)(a+b+1)\Gamma(a+b)\Gamma(a)} - \left(\frac{a}{a+b}\right)^2 \\&= \frac{a(a+1)}{(a+b)(a+b+1)} - \left(\frac{a}{a+b}\right)^2 \\&= \frac{ab}{(a+b)^2(a+b+1)},\end{aligned}$$

where, again, in the fifth step, we used the fact that the gamma distribution is normalized, and in the sixth step, we used the property  $\Gamma(x+1) = x\Gamma(x)$ .