

**Theorem** (uniqueness; without proof).

If  $M(t)$  is a mgf, finite on  $(-\delta, \delta)$ , then there is exactly one distribution that  $M(t)$  is the mgf for.

**Theorem.** If  $M_X(t)$  is the mgf for  $X$  and finite on  $(-\delta, \delta)$ , then every moment of  $X$  is finite and

$$E(X^m) = \frac{d^m}{dt^m} M_X(t) \Big|_{t=0} \quad \text{for each } m = 1, 2, \dots$$

The sequence of moments is unique (i.e. not shared by other distributions).

**Sketch of proof.** By our assumptions on  $M_X$  we can write  $\frac{d^m}{dt^m} E(e^{tX}) = E(\frac{d^m}{dt^m} e^{tX}) = E(\frac{d^{m-1}}{dt^{m-1}} X e^{tX}) = \dots = E(X^m e^{tX})$  which is  $E(X^m)$  for  $t = 0$ . The sequence of moments is unique since the Taylor coefficients of  $M_X(t)$  are unique: consider  $M_X(t) = E(e^{tX}) = E(\sum_{n=0}^{\infty} \frac{(tX)^n}{n!}) = \sum_{n=0}^{\infty} \frac{E(X^n)}{n!} t^n$ .  $\square$

**Example** (cont.). For  $Z \sim N(0, 1)$  we have  $M_Z(t) = e^{t^2/2}$  and therefore  $E(Z) = \frac{d}{dt} M_Z(t) \Big|_{t=0} = te^{t^2/2} \Big|_{t=0} = 0$  (which we already know). Similarly, by applying the product rule, one can compute higher moments, e.g.  $E(Z^2) = \frac{d^2}{dt^2} M_Z(t) \Big|_{t=0} = \frac{d}{dt} te^{t^2/2} \Big|_{t=0} = \dots = 1$ ,  $E(Z^3) = 0$ ,  $E(Z^4) = 3$ .

**Example.** Consider  $X \sim \text{Poisson}(\lambda)$ ,  $\lambda > 0$ , i.e.  $f_X(x) = \lambda^x e^{-\lambda}/x!$ ,  $x = 0, 1, 2, \dots$ . As in the previous “sketch of proof” use the Taylor expansion of the exponential function to obtain  $M_X(t) = E(e^{tX}) = \sum_{x=0}^{\infty} e^{tx} \lambda^x e^{-\lambda}/x! = e^{-\lambda} \sum_{x=0}^{\infty} (e^t \lambda)^x / x! = e^{-\lambda} e^{e^t \lambda} = e^{\lambda(e^t - 1)}$ . Therefore  $E(X) = \frac{d}{dt} M_X(t) \Big|_{t=0} = e^{\lambda(e^t - 1)} \lambda e^t \Big|_{t=0} = \lambda$  and, similarly,  $E(X^2) = \lambda(\lambda + 1)$ . This gives  $\text{var}(X) = E(X^2) - \{E(X)\}^2 = \lambda$ .

**Theorem** (without proof). Let  $X, X_1, X_2, \dots$  be random variables and suppose that all mgf’s are finite on  $(-\delta, \delta)$  for some  $\delta > 0$ . Then, as  $n \rightarrow \infty$ ,

$$\begin{aligned} M_{X_n}(t) &\rightarrow M_X(t) \quad \text{for all } t \in (-\delta, \delta) \\ \iff F_{X_n}(x) &\rightarrow F_X(x) \quad \text{for all } x \text{ where } F_X \text{ is continuous,} \end{aligned}$$

i.e.  $X_n$  converges in distribution to  $X$ .

**Example** (Poisson approximation of the binomial distribution).

Consider  $Y_n \sim \text{binomial}(n, p)$  where  $np_n \rightarrow \lambda > 0$  as  $n \rightarrow \infty$  (and thus  $p_n \rightarrow 0$ ). Then  $Y_n$  converges in distribution to a Poisson distribution:  $M_{Y_n}(t) = (p_n e^t + 1 - p_n)^n = (1 + \frac{(e^t - 1)(np_n - \lambda + \lambda)}{n})^n = (1 + \frac{(e^t - 1)\lambda + \varepsilon_n}{n})^n$  with  $\varepsilon_n \rightarrow 0$  and therefore  $\lim_{n \rightarrow \infty} (1 + \frac{(e^t - 1)\lambda + \varepsilon_n}{n})^n = e^{\lambda(e^t - 1)}$  which is the mgf of the Poisson distribution. This result allows us to use the Poisson distribution to approximate binomial probabilities, provided  $n$  is large and  $p$  very small. This is illustrated by the following numerical example.

Suppose the incidence of fatality from a certain vaccination is 1 in 5,000. After vaccinating 10,000 people there are five fatalities (although “only” two are expected). Should you be concerned?

A better question here would be: If  $p = P(\text{“success”}) = 0.0002$  is true, is seeing 5 or more fatalities unexpected? (Is the probability for this event high or low?) We can answer this question by using the Poisson approximation: the number of successes  $Y_n$  is binomially distributed with parameter  $p = 0.0002$  (which is small) and  $n = 10,000$  (which is large). Approximate  $Y_n$  by  $X \sim \text{Poisson}(\lambda)$  with  $\lambda \approx np = 2$ :  $P(Y_n \geq 5) \approx P(X \geq 5) = 1 - P(X < 5) = 1 - \sum_{x=0}^4 e^{-2} 2^x / x! \approx 0.05 = 5\%$ .

### 3 Families of distributions

#### 3.1 Occurrences and waiting time

**Example** (unfair coin).

$$P(\text{first H on } n\text{-th flip}) = p(1-p)^{n-1},$$

$$P(k\text{-th H on } n\text{-th flip}) = P(X = n) = \binom{n-1}{k-1} p^k (1-p)^{n-k}, \quad n \geq k.$$

Now set  $Y = \#$  failures until the  $k$ -th success occurs. Then  $Y = X - k$  and  $f_Y(y)$  is the pmf of the **negative binomial** distribution, i.e. for  $y = 0, 1, 2, \dots$  holds

$$\begin{aligned} f_Y(y) &= P(Y = y) = P(X - k = y) = P(X = y + k) = f_X(y + k) \\ &= \binom{y + k - 1}{k - 1} p^k (1 - p)^{y + k - k}. \end{aligned}$$

*Recall:* In a sequence of independent Bernoulli trials the **binomial** distribution is the distribution of the number of successes in  $n$  trials; the **geometric** distribution is the distribution of the number of trials until the first success occurs.

**Theorem** The geometric distribution is **memoryless**, i.e. if  $X \sim \text{geometric}(p)$ , then for  $t, x \in \{0, 1, 2, \dots\}$

$$P(X > t + x | X > t) = P(X > x) \quad [= P(\text{"wait } x \text{ trials at } t = 0\text{"})].$$

**Proof.** Let  $A_i = \text{"success on } i\text{-th trial"}$  and  $X = \#$  trials until the first success occurs,  $X \sim \text{geometric}(p)$ . Then

$$\text{"}X > t\text{"} = A_1^c \cap \dots \cap A_t^c, \quad \text{"}X > t + x\text{"} = A_1^c \cap \dots \cap A_{t+x}^c$$

and, using the independence of the events  $A_1, A_2, \dots$ ,

$$P(X > t + x) = P(A_1^c \dots A_t^c) P(A_{t+1}^c \dots A_{t+x}^c) = P(X > t) P(X > x).$$

The last equality holds since  $P(A_{t+1}^c \dots A_{t+x}^c) = P(A_1^c \dots A_x^c) = P(X > x)$ . Therefore,

$$P(X > t + x | X > t) = \frac{P(X > t + x)}{P(X > t)} = \frac{P(X > t) P(X > x)}{P(X > t)} = P(X > x). \quad \square$$

**Remark.** We can use probabilistic arguments to derive the *mean of the negative binomial distribution*:

$\#$  trials to the first success  $\sim \text{geometric}(p)$  (with mean  $1/p$ )

$\Rightarrow$  the distribution of " $\#$  trials to the *next* success" has mean  $1/p$

$\Rightarrow E(\text{"}\# \text{ failures to the } k\text{-th success"})$

$$= E(\text{"}\# \text{ trials to the } k\text{-th success"}) - \# \text{ successes} = k/p - k = k(1-p)/p$$

(= mean of the negative binomial distribution).

**Theorem.** Let  $N_k \sim \text{negative binomial}(k, p)$ ,  $B_n \sim \text{binomial}(n, p)$ , then

$$P(N_k + k \leq n) = P(B_n \geq k).$$

**Proof.** " $N_k \leq n - k$ " = "at least  $k$  successes in  $n$  trials" = " $B_n \geq k$ ".  $\square$