nurdan.kar@bilkent.edu.tr

3.5. The binomial probability distributions

Definition. A random variable Y possess a binomial distribution if

1. The experiment consists of a fixed number n of trials.
2. Each trial can result in one of only two possible outcomes, called "success" and "failure".
3. The probability of "success" p, is constant from trial to trial.
4. The trials are independent.

An experiment for which Conditions 1-4 are satisfied is called a binomial experiment.

Example. Suppose a state has 500,000 licensed drivers, of whom 400,000 are insured. A sample of ten drivers is chosen without replacement. The ith trial is labeled S if the ith driver chosen is insured. Can the distribution of the number of insured drivers in the sample be modeled by a Binomial distribution?

Solution. 1. The number of trials is fixed: $n=10$.

2. Each trial has two outcomes: insured (success), otherwise (failure)

3. $P(S \text{ on } 2 \mid S \text{ on } 1) = \dfrac{399,999}{499,999} = .7999996 = .8$

$P(S \text{ on } 10 \mid S \text{ on first } 9) = \dfrac{399,991}{499,991} = .799996 \approx .80000$

$P(S \text{ on } 2) = \dfrac{400,000}{500,000} = .8.$

$P(S \text{ on } 2 \mid S \text{ on } 1) \neq P(S \text{ on } 2)$. Even though the values are very close, they are not exactly the same. Therefore, they are not exactly independent. However, for practical purposes it can be treated as independent.

4. The probability of "success" is $P(S) = .8$.

Conclusion: The experiment is binomial with $n=10$ and $p=.8$.

In general, if sampling is without replacement, the experiment will not yield independent trials. However, we will use the following rule in deciding whether a "without replacement" experiment can be treated as a binomial experiment.

Rule: In sampling without replacement, if the sample size (number of trials), n, is at most 5% of the population size, N, the experiment can be treated as a binomial experiment.

* In previous example, $\frac{n}{N} = \frac{10}{500,000} < .05$. Thus, it is a Binomial experiment.

The binomial random variable and distribution

Definition. Given a binomial experiment consisting of n trials, the binomial random variable X associated with this experiment is defined as

$$X = \text{the number of successes among the } n \text{ trials}$$

NOTATION.
* $X \sim Bin(n,p)$: X is a binomial rv based on n trials with success probability, p.
* $b(x; n,p)$: pmf of a binomial rv X with parameters n and p.

Theorem.

$$b(x;n,p) = \begin{cases} \binom{n}{x} p^x (1-p)^{n-x} & , x = 0,1,\ldots,n \\ 0 & , \text{otherwise} \end{cases}$$

Example. A restaurant serves 8 entrées of fish, 12 of beef, and 10 of poultry. If customers select from these entrées randomly, what is the probability that two of the next four customers order fish entrées?

Solution. X: the number of fish entrées (successes) ordered by the next four customers

n: four customers

p: success probability $(8/30 = 4/15)$

Then, $b(2;4,4/15) = \binom{4}{2}\left(\frac{4}{15}\right)^2\left(1-\frac{4}{15}\right)^2 = .23$.

Computing binomial probabilities

NOTATION. For $X \sim Bin(n,p)$, the cdf will be denoted by

$$B(x;n,p) = P(X \le x) = \sum_{y=0}^{x} b(y;n,p), \quad x = 0,1,\dots,n.$$

Example. Suppose that 20% of all copies of a particular textbook fail a binding strength test. Let X denote the number among 15 randomly selected copies that fail the test. Then, X has a binomial distribution with $n=15$ and $p=.2$; that is $Bin(15,.2)$. Find

a) the probability that at most 8 fail the test

b) the probability that between 4 and 7, inclusive, fail the test.

Solution. a) We wish to find $P(X \le 8)$.

$$P(X \le 8) = \sum_{y=0}^{8} b(y;15,.2) = B(8;15,.2) = .999$$ (From binomial cdf table)

b) We wish to find $P(4 \le X \le 7)$.

$$\begin{aligned} P(X=4, \text{ or } 5, \text{ or } 6, \text{ or } 7) &= P(X \le 7) - P(X \le 3) \\ &= \sum_{y=0}^{7} b(y;15,.2) - \sum_{y=0}^{3} b(y;15,.2) \\ &= B(7;15,.2) - B(3;15,.2) \\ &= .996 - .648 \quad \text{(from cdf (binomial) table)} \\ &= .348 \end{aligned}$$

The mean and variance of a binomial random variable

Proposition. If $X \sim Bin(x,p)$, then $E(X) = np$, $V(X) = np(1-p)$, and $\sigma_X = \sqrt{np(1-p)}$.

An important use of binomial distribution is to estimating the precision of simulated probabilities. Consider the relative frequency definition of probability, $\hat{P}(A) = X/n$, where n is the number of trials, and X is the number of trials in which event A occurred. Then,

$$E[\hat{P}(A)] = E\left[\frac{1}{n} \cdot X\right] \underset{\text{by the property of the exp. value}}{=} \frac{1}{n} \cdot E(X) \underset{\text{by the previous proposition}}{=} \frac{1}{n} \cdot np = p = P(A)$$

and

$$\sigma_{\hat{P}(A)} = \sigma_{X/n} = \left|\frac{1}{n}\right| \cdot \sigma_X = \frac{1}{n}\sqrt{np(1-p)} = \sqrt{\frac{np(1-p)}{n^2}} = \sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{P(A)[1-P(A)]}{n}}.$$

by the previous proposition

The expression above is called the <u>standard error</u> of $\hat{P}(A)$. It measures the typical deviation of $\hat{P}(A)$ from $P(A)$. However, this formula is not handy as we usually simulate a probability when $P(A)$ is unknown. Therefore, we substitute the estimate of a probability, $\hat{P}(A)$, in the formula above and obtain the <u>estimated standard error formula</u>:

$$\sigma_{\hat{P}(A)} \approx \sqrt{\frac{\hat{P}(1-\hat{P})}{n}}.$$

<u>The mgf of a binomial rv</u>

In order to determine the mgf of a binomial rv, we use the binomial theorem, $(a+b)^n = \sum_{x=0}^{n} \binom{n}{x} a^x b^{n-x}$, (*)

and the definition of the mgf

$$M_X(t) = E(e^{tX}) = \sum_{x \in D} e^{tx} p(x) = \sum_{x=0}^{n} e^{tx} \binom{n}{x} p^x (1-p)^{n-x} = \sum_{x=0}^{n} \binom{n}{x} (pe^t)^x (1-p)^{n-x} = (pe^t + 1 - p)^n.$$

by the def. of binomial pmf

by the binomial theorem given by (*).

The necessary condition of the mgf, $M_X(0) = 1$, is also valid here. We can derive the mean and the variance by differentiating the mgf, $M_X(t)$.

First derivative yields

$$M_X'(t) = n(pe^t + 1 - p)^{n-1} \cdot pe^t \Rightarrow M_X'(0) = np = \mu = E(X).$$

Second derivative yields

$$M_X''(t) = n(n-1)(pe^t + 1 - p)^{n-2} \cdot pe^t \cdot pe^t + n(pe^t + 1 - p) \cdot pe^t$$

$$\Rightarrow M_X''(0) = n(n-1)p^2 + np = E(X^2).$$

Thus, we have

$$V(X) = E(X^2) - [E(X)]^2 = n(n-1)p^2 + np - n^2p^2 = np - np^2 = np(1-p).$$

$$\cancel{n^2p^2} - np^2 + np - \cancel{n^2p^2}$$

Example. Two proofreaders, Ruby and Myra, read a book independently and found r and m misprints, respectively. Suppose that the probability that a misprint is noticed by Ruby is p and the probability that it is noticed by Myra is q, where these two probabilities are independent. If the number of misprints noticed by both Ruby and Myra is b, estimate the number of unnoticed misprints. [The problem was posed and solved by George Pólya in Jan 1976.]

Solution. * M: the total number of misprints in the book.

* The expected number of misprints that may be noticed by Ruby, Myra, and both of them are Mp, Mq, and Mpq, respectively.

We assume that expected number of misprints approximates the mistakes found. Namely, $E[X_R] = Mp \approx r$, $E[X_M] = Mq \approx m$, and $E[X_{both}] = Mpq \approx b$.

From $Mpq \approx b$, we write

$$M \approx \frac{b}{pq} \Rightarrow M \approx \frac{b}{(r/M)(m/M)} \Rightarrow M \approx \frac{M^2 b}{rm} \Rightarrow M \approx \frac{rm}{b},$$

which is the number of total misprints. Then, the number of unnoticed misprints is estimated by

$$M-(r+m-b) \approx \frac{rm}{b} - (r+m-b) = \frac{rm-br+bm-b^2}{b} = \frac{r(m-b)+b(m-b)}{b}$$

$$= \frac{(m-b)(r+b)}{b}.$$

Here, each misprint is a Bernoulli experiment.

That is, $p(\text{noticed}) = 1$ and $p(\text{not noticed}) = 0$.

Definitions of the Mp, Mq, and Mpq come from the first derivative of the mgf of a binomial rv.

3.6. The Poisson probability distribution

In many natural phenomena and real-world problems, we encounter binomial rv. However, the direct calculation of the binomial pmf can be computationally burdensome. In 1837, French mathematician Simeon-Denis Poisson introduced an approximation to the binomial pmf.

Definition. A rv X is said to have a Poisson distribution with parameter μ $(\mu > 0)$ if the pmf of X is

$$p(x;\mu) = \frac{e^{-\mu}\mu^x}{x!}, \quad x = 0,1,2,\ldots$$

Here, μ is the expected value of X.

The Poisson distribution meets two conditions of pmf:

1. $p(x;\mu) > 0$
2. $$\sum_{x=0}^{\infty} \frac{e^{-\mu}\mu^x}{x!} = e^{-\mu} \underbrace{\sum_{x=0}^{\infty} \frac{\mu^x}{x!}}_{\text{Taylor series expansion of } e^{\mu}} = e^{-\mu} e^{\mu} = 1.$$

Example. The atoms of a radioactive element are randomly disintegrating. If every gram of this element, on average, emits 3.9 alpha particles per second, what is the probability that during the next second the number of alpha particles emitted from 1 gram is

a) at most 6 b) at least 2 c) at least 3 and at most 6 ?

Solution. • X = total number of atoms that emit an alpha particle in the next second.

• Each atom is a binomial trial; success: it emits an alpha particle in the next second
failure: it does not

• $E(X) = \mu = 3.9 = np \Rightarrow p = 3.9/n$

Since n is very large (every gram consists of many atoms), p is very small, and hence, to a very close approximation, X has a Poisson distribution with the parameter $\mu = 3.9$. Then, $p(n; 3.9) = \dfrac{(3.9)^n e^{-3.9}}{n!} = P(X=n)$.

NEXT PAGE

a) $P(X \leq 6) = \sum_{n=0}^{6} \frac{(3.9)^n \cdot e^{-3.9}}{n!} \approx .899$

b) $P(X \geq 2) = 1 - P(X \leq 2) = 1 - P(X=0 \text{ or } 1) = 1 - \frac{(3.9)^0 e^{-3.9}}{0!} - \frac{(3.9)^1 e^{-3.9}}{1!} = .901.$

c) $P(3 \leq X \leq 6) = \sum_{n=3}^{6} \frac{(3.9)^n e^{-3.9}}{n!} \approx .646$

The Poisson distribution as a limit

Proposition. Suppose that in the binomial pmf $b(x;n,p)$ we let $n \to \infty$ and $p \to 0$ in such a way that np approaches a value $\mu > 0$. Then, $b(x;n,p) \to p(x;\mu)$.

Proof. Let us consider the definition of binomial pmf

$$b(x;n,p) = \binom{n}{x} p^x (1-p)^{n-x}$$

$$= \frac{n!}{(n-x)!\,x!} p^x (1-p)^{n-x}$$

$$= \frac{n!}{(n-x)!\,x!} p^x (1-p)^{n-x} \cdot \frac{n^x}{n^x}$$

$\frac{n(n-1)\dots(n-x+1)(n-x)!}{(n-x)!\,x!}$

can be written as $\left(1-\frac{x}{n}\right)$

$$= \frac{n(n-1)\dots(n-x+1)}{n \cdot n \dots n} \frac{(np)^x}{x!} \frac{(1-p)^n}{(1-p)^x}$$

Taking the limit as $n \to \infty$ and $p \to 0$ with $np \to \mu$, we obtain

$$\lim_{n\to\infty} (x;n,p) = \underbrace{\lim_{n\to\infty}\left(1-\frac{x}{n}\right)}_{1} \cdot \frac{(np)^x}{x!} \cdot \underbrace{\lim_{n\to\infty}\left(1-\frac{\mu}{n}\right)^n}_{e^{-\mu}} \cdot \underbrace{\lim_{n\to\infty}\frac{1}{(1-\mu/n)^x}}_{1}$$

$$= \frac{(np)^x}{x!} e^{-\mu}$$

$$= \frac{\mu^x e^{-\mu}}{x!}$$

$$= p(x;\mu). \blacksquare$$

The mean, variance, and mgf of a Poisson random variable

Proposition. If X has a Poisson distribution with parameter μ, then $E(X) = V(X) = \mu$.

Proposition. The Poisson mgf is $M_X(t) = e^{\mu(e^t-1)}$.

Proof. By the definition of mgf,

$$M_X(t) = E(e^{tX}) = \sum_{x=0}^{\infty} e^{tx} \underbrace{e^{-\mu}\frac{\mu^x}{x!}}_{p(x;\mu)} = e^{-\mu} \underbrace{\sum_{x=0}^{\infty} \frac{(\mu e^t)^x}{x!}}_{\substack{\text{Taylor series} \\ \text{expansion of } \mu e^t}} = e^{-\mu} e^{\mu e^t} = e^{\mu(e^t-1)}. \blacksquare$$

Example. Every week the average number of wrong-number phone calls received by a certain mail-order house is seven. What is the probability that they will receive a) two wrong calls tomorrow;

b) at least one wrong call tomorrow?

Solution. X: the number of wrong numbers received tomorrow (Poisson r.v.)

$\mu = E(X) = 1$

Thus, we have

$$P(X=n) = \frac{e^{-1}\cdot(1)^n}{n!} = \frac{1}{e\cdot n!}.$$

a) $P(X=2) = \frac{1}{2e} \approx .18$, b) $P(X\geq 1) = 1 - P(X=0) = 1 - \frac{1}{e} \approx .63$.

The Poisson process

The Poisson distribution appears in connection with the study of sequences of random events occurring over time. Suppose that, starting from a time point labeled $t=0$, we start counting the number of events. Then for each value of t, we obtain a number denoted by n, which is the number of events that have occurred during $[0,t]$.

Here, we consider an arrival process with any real number t is a possible arrival time.

We define

$$P(n,\tau) = P(\text{there are exactly } n \text{ arrivals during an interval of length } \tau),$$

and assume that this probability is the same for all intervals of the same length τ. We also introduce $\lambda > 0$ for the arrival rate or intensity of the process.

Definition. An arrival process is called a Poisson process with rate λ if it has the following properties:

a. Time homogeneity. The probability $P(n,\tau)$ of n arrivals is the same for all intervals of the same length τ.

b. Independence. The number of arrivals during a particular interval is independent of the history of arrivals outside this interval.

c. Small interval probabilities. The probabilities $P(n,\tau)$ satisfy

$$P(0,\tau) = 1-\lambda\tau + o(\tau), \quad \rightarrow \text{prob. of no arrivals}$$

$$P(1,\tau) = \lambda\tau + o_1(\tau). \quad \rightarrow \text{prob. of one arrival}$$

$o(\tau)$ and $o_1(\tau)$ represent the tiny prob. of two or more arrivals.

Here $o(\tau)$ and $o_1(\tau)$ are functions of τ that satisfy

$$\lim_{\tau\to 0}\frac{o(\tau)}{\tau} = 0 \quad \text{and} \quad \lim_{\tau\to 0}\frac{o_1(\tau)}{\tau} = 0.$$

The first property says that arrivals are equally likely at all times. This property is a counterpart of the success probability, p, in a Bernoulli process is constant over time.

The second property is analogous to the independence of trials in a Bernoulli process. Consider that for all $n \geq 0$, and for any time interval $(t, t+s)$, the probability of n events in $(t, t+s)$ is independent of how many events have occurred earlier or how they have occured. In other words, suppose that we are given complete or partial information on the arrivals outside $(t, t+s)$. This property says that the given information is irrelevant.

The third property says that the occurrence of two or more arrivals in a very small time interval is practically impossible. This is given by the limits in the property. This implies that as $\tau \to 0$, the probability of two or more arrivals, $o(\tau)$, $o_1(\tau)$ approaches 0 faster than τ does. That is, if τ negligible, then $o(\tau)$ and $o_1(\tau)$ are even more negligible. Note that the prob. of two or more arrivals is $1-P(0,\tau)-P(1,\tau) = -o(\tau)-o_1(\tau)$.

$\nearrow$ E[arrivals] $= \lambda\tau$, V(arrivals) $= \lambda\tau$.

Proposition. $P(n,\tau) = \frac{e^{-\lambda\tau}(\lambda\tau)^n}{n!}$, so that the number of arrivals during a time interval of length τ is a Poisson rv with parameter $\mu = \lambda\tau$. The expected number of arrivals during any such time interval is then $\lambda\tau$, so the expected number during a unit interval of time is λ.

In the definition of the Poisson process, if we use "event" instead of "arrival", then the number of events occuring during a fixed time interval of length t has a Poisson distribution with parameter λt.

Example. Let $N(t)$ be the number of earthquakes that occur at or prior to time t worldwide. Suppose that $\{N(t): t \geq 0\}$ is a Poisson process with rate λ and the probability that the magnitude of an earthquake on the Richter scale is 5 or more is p. Find the probability of k earthquakes of such magnitudes at or prior to t worldwide.

Solution. Let $X(t)$ be the number of earthquakes of magnitude 5 or more on the Ritcher scale at or prior to t worldwide. We have the sequence of events $\{N(t)=n\}$, $n=0,1,2,\ldots$, which meets $\{N(t)=n\} \cap \{N(t)=m\} = \phi$ for all $n \neq m$ and $\bigcup_{n=0}^{\infty} \{N(t)=n\} = \Omega$. Then, by the law of total probability,

$$P(X(t)=k) = \sum_{n=0}^{\infty} P(X(t)=k \mid N(t)=n) \cdot P(N(t)=n).$$

When $n<k$, we get $P(X(t) \mid N(t)=n) = 0$. When $n \geq k$, $P(X(t)=k \mid N(t)=n)$ is binomial with fixed number of trials (n) and success (magnitude ≥ 5) prob., p.

Then,

$$P(X(t)=k) = \sum_{n=k}^{\infty} \binom{n}{k} p^k (1-p)^{n-k} \frac{e^{-\lambda t}(\lambda t)^n}{n!}$$

$$= \sum_{n=k}^{\infty} \frac{\cancel{n!}}{k!(n-k)!} p^k (1-p)^{n-k} \frac{e^{-\lambda t}(\lambda t)^k (\lambda t)^{n-k}}{\cancel{n!}}$$

$$= \sum_{n=k}^{\infty} \frac{1}{k!(n-k)!} (\lambda t p)^k e^{-\lambda t} [(1-p)(\lambda t)]^{n-k}$$

$$= \frac{1}{k!} (\lambda t p)^k e^{-\lambda t} \sum_{n=k}^{\infty} \frac{1}{(n-k)!} [(1-p)(\lambda t)]^{n-k}$$

$$= \frac{(\lambda t p)^k e^{-\lambda t}}{k!} \cdot \sum_{j=0}^{\infty} \frac{1}{j!} [(1-p)\lambda t]^j = \frac{(\lambda t p)^k e^{-\lambda t} \cdot e^{\lambda t (1-p)}}{k!} = \frac{(\lambda t p)^k e^{-p\lambda t}}{k!}.$$

Taylor expansion of $(1-p)\lambda t$

$$\sum_{j=0}^{\infty} \frac{1}{j!} [(1-p)\lambda t]^j$$

Poisson dist. with rate $\lambda t p$.

Conclusion. $\{X(t): t \geq 0\}$ is a Poisson process with parameter λp.