

Econ 205

Discrete Random Variables and Probability Distributions – 2

Course Outline

- **Probability**

- Probability and Inference
- The Discrete Case
- Calculating the Probability of an Event
- Conditional Probability and the Independence of Events
- Laws of Probability
- The Law of Total Probability and Bayes' Rule
- Introduction to Random Variables and their distributions

- **Discrete Random Variables and Their Probability Distributions**

- Discrete Random Variable
- The Probability Distribution for Discrete Random Variable
- The Expected Value of Random Variable
- Moments and Moment-Generating Functions
- The Bernoulli Probability Distribution
- The Binomial Probability Distribution
- The Geometric Probability Distribution
- The Negative Binomial Probability Distribution
- The Hypergeometric Probability Distribution
- The Poisson Probability Distribution
- Probability-Generating Functions

- **Continuous Random Variables and Their Probability Densities**

- Continuous Random Variable
- The Probability Distribution for Continuous Random Variable
- The Expected Value for Continuous Random Variable
- Moments and Moment-Generating Functions
- The Uniform Probability Distribution
- The Normal Probability Distribution
- Exponential Probability Distribution
- Chi-Square Probability Distribution
- Tchebysheff's Theorem

Course Outline (cont'd)

- **Multivariate Probability Distributions**

- Bivariate and Multivariate Probability Distributions
- Independent Random Variables
- Marginal and Conditional Probability Distributions
- The Expected Value of a Function of Random Variables
- Special Theorems
- The Covariance of Two Random Variables
- The Expected Value and Variance of Linear Functions of Random Variables
- The Multinomial Probability Distribution
- The Bivariate Normal Distribution
- Conditional Expectations

- **Functions of Random Variables**

- Finding the Probability Distribution of a Function of Random Variables
- The Method of Distribution Functions
- The Method of Transformations
- The Method of Moment-Generating Functions
- Multivariable Transformations Using Jacobians

Table of Contents of Devore & Berk

2	Probability	50
	Introduction	50
2.1	Sample Spaces and Events	51
2.2	Axioms, Interpretations, and Properties of Probability	56
2.3	Counting Techniques	66
2.4	Conditional Probability	74
2.5	Independence	84
3	Discrete Random Variables and Probability Distributions	96
	Introduction	96
3.1	Random Variables	97
3.2	Probability Distributions for Discrete Random Variables	101
3.3	Expected Values of Discrete Random Variables	112
3.4	Moments and Moment Generating Functions	121
3.5	The Binomial Probability Distribution	128
3.6	Hypergeometric and Negative Binomial Distributions	138
3.7	The Poisson Probability Distribution	146
4	Continuous Random Variables and Probability Distributions	158
	Introduction	158
4.1	Probability Density Functions and Cumulative Distribution Functions	159
4.2	Expected Values and Moment Generating Functions	171
4.3	The Normal Distribution	179
4.4	The Gamma Distribution and Its Relatives	194
4.5	Other Continuous Distributions	202
4.6	Probability Plots	210
4.7	Transformations of a Random Variable	220
5	Joint Probability Distributions	232
	Introduction	232
5.1	Jointly Distributed Random Variables	233
5.2	Expected Values, Covariance, and Correlation	245
5.3	Conditional Distributions	253
5.4	Transformations of Random Variables	265
5.5	Order Statistics	271

Content of this presentation

- Discrete Random Variables and Probability Distributions
 - The Binomial Probability Distribution
 - The Geometric Probability Distribution
 - The Negative Binomial Probability Distribution
 - The Hypergeometric Probability Distribution
 - The Poisson Probability Distribution
- Examples

The Binomial and Negative Binomial Probability Distributions

Binomial experiment:

- 1) The experiment consists of a sequence of n smaller experiments called **trials**, where n is fixed in advance of the experiment.
- 2) Each trial can result in one of the same two possible outcomes, which we denote by **success** (S) or **failure** (F).
- 3) The trials **are independent**.
- 4) The probability of success (**p**) is **constant** from trial to trial.

The **binomial random variable X** associated with this experiment is defined as

X = the number of S's among the n trials.

Because the pmf of a binomial rv X depends on the two parameters n and p , we denote the pmf by $b(x; n, p)$.

$$b(x; n, p) = \left\{ \begin{array}{l} \text{number of sequences of} \\ \text{length } n \text{ consisting of } x \text{ S's} \end{array} \right\} \cdot \left\{ \begin{array}{l} \text{probability of any} \\ \text{particular such sequence} \end{array} \right\}$$

$$b(x; n, p) = \begin{cases} \binom{n}{x} p^x (1-p)^{n-x} & x = 0, 1, 2, \dots, n \\ 0 & \text{otherwise} \end{cases}$$

The **Negative Binomial random variable and distribution are based on an experiment satisfying the following:**

- 1) The experiment consists of a sequence of **trials**
- 2) Each trial can result in one of the same two possible outcomes, which we denote by **success** (S) or **failure** (F).
- 3) The trials **are independent**.
- 4) The probability of success (**p**) is **constant** from trial to trial. So $P(\text{S on trial } i) = p$.
- 5) The experiment continues (trials are performed) until a total of r successes has been observed, where r is a specified positive integer.

The **negative binomial random variable X** associated with this experiment is defined as

X = the number of F's that precede the r^{th} S.

Possible values of X are $0, 1, 2, \dots$. Let $nb(x; r, p)$ denote the pmf of X . The event $\{X = x\}$ is equivalent to $\{r - 1 \text{ S's in the first } (x + r - 1) \text{ trials and an S on the } (x + r)\text{th trial}\}$.

The pmf of the negative binomial rv X with parameters $r = \text{number of S's}$ and $p = P(S)$ is

$$nb(x; r, p) = \binom{x + r - 1}{r - 1} p^r (1 - p)^x \quad x = 0, 1, 2, \dots$$

The Geometric Probability Distribution:

Consider the pmf of the negative binomial probability distribution.

In the special case $r = 1$, the pmf is

$$nb(x; 1, p) = (1 - p)^x p \quad x = 0, 1, 2, \dots \quad (3.18)$$

Both $X = \text{number of } F\text{'s}$ and $Y = \text{number of trials } (= 1 + X)$ are referred to in the literature as **geometric random variables**, and the pmf in (3.18) is called the **geometric distribution**. The name is appropriate because the probabilities form a geometric series: $p, (1 - p)p, (1 - p)^2 p, \dots$. To see that the sum of the probabilities is 1, recall that the sum of a geometric series is $a + ar + ar^2 + \dots = a/(1 - r)$ if $|r| < 1$, so for $p > 0$,

$$p + (1 - p)p + (1 - p)^2 p + \dots = \frac{p}{1 - (1 - p)} = 1$$

Note that we need to calculate the mean, variance and the mgf of the geometric distribution differently depending on whether we define it as X or Y .

The Binomial Probability Distribution: The mean, variance and mgf

We will use the following rule of thumb in deciding whether a “without-replacement” experiment can be treated as a binomial experiment.

Consider sampling without replacement from a dichotomous population of size N . If the sample size (number of trials) n is at most 5% of the population size, the experiment can be analyzed as though it were exactly a binomial experiment.

We will often write $X \sim \text{Bin}(n, p)$ to indicate that X is a binomial rv based on n trials with success probability p .

For $X \sim \text{Bin}(n, p)$, the cdf will be denoted by

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

The Mean and Variance of X :

Proposition:

If $X \sim \text{Bin}(n, p)$, then $E(X) = np$, $V(X) = np(1-p) = npq$, and $\sigma_X = \sqrt{npq}$ (where $q = 1 - p$).

The **Moment Generating Function** of X and using the mgf to obtain the Mean and the Variance:

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

Here we have used the binomial theorem, $\sum_{x=0}^n a^x b^{n-x} = (a+b)^n$.

Notice that the mgf satisfies the property required of all moment generating functions, $M_X(0) = 1$, because the sum of the probabilities is 1. The mean and variance can be obtained by differentiating $M_X(t)$:

$$M'_X(t) = n(pe^t + 1 - p)^{n-1} pe^t \text{ and } \mu = M'_X(0) = np$$

Then the second derivative is

$$M''_X(t) = n(n-1)(pe^t + 1 - p)^{n-2} pe^t pe^t + n(pe^t + 1 - p)^{n-1} pe^t$$

and

$$E(X^2) = M''_X(0) = n(n-1)p^2 + np$$

Therefore,

$$\begin{aligned} \sigma^2 &= V(X) = E(X^2) - [E(X)]^2 \\ &= n(n-1)p^2 + np - n^2p^2 = np - np^2 = np(1-p) \end{aligned}$$

There is a missing term here: $\binom{n}{x}$

The Negative Binomial Probability Distribution: The mean, variance and mgf

Proposition:

If X is a negative binomial rv with pmf $nb(x; r, p)$, then

$$M_X(t) = \frac{p^r}{[1 - e^t(1-p)]^r} \quad E(X) = \frac{r(1-p)}{p} \quad V(X) = \frac{r(1-p)}{p^2}$$

Proof In order to derive the moment generating function, we will use the binomial theorem as generalized by Isaac Newton to allow negative exponents, and this will help to explain the name of the distribution. If n is any real number, not necessarily a positive integer,

$$(a + b)^n = \sum_{x=0}^{\infty} \binom{n}{x} b^x a^{n-x}$$

where

$$\binom{n}{x} = \frac{n \cdot (n-1) \cdot \dots \cdot (n-x+1)}{x!} \quad \text{except that } \binom{n}{0} = 1$$

In the special case that $x > 0$ and n is a negative integer, $n = -r$,

$$\begin{aligned} \binom{-r}{x} &= \frac{-r \cdot (-r-1) \cdot \dots \cdot (-r-x+1)}{x!} \\ &= \frac{(r+x-1)(r+x-2) \cdot \dots \cdot r}{x!} (-1)^x = \binom{r+x-1}{r-1} (-1)^x \end{aligned}$$

Using this in the generalized binomial theorem with $a = 1$ and $b = -u$,

$$(1-u)^{-r} = \sum_{x=0}^{\infty} \binom{r+x-1}{r-1} (-1)^x (-u)^x = \sum_{x=0}^{\infty} \binom{r+x-1}{r-1} u^x$$

Now we can find the moment generating function for the negative binomial distribution:

$$\begin{aligned} M_X(t) &= \sum_{x=0}^{\infty} e^{tx} \binom{r+x-1}{r-1} p^r (1-p)^x = p^r \sum_{x=0}^{\infty} \binom{r+x-1}{r-1} [e^t(1-p)]^x \\ &= \frac{p^r}{[1 - e^t(1-p)]^r} \end{aligned}$$

The mean and variance of X can now be obtained from the moment generating function. ■

$$(1-u)^{-r} = \sum_{x=0}^{\infty} \binom{-r}{x} (-u)^x 1^{-r-x}$$

The Negative Binomial Probability Distribution: Alternative definition

The Negative Binomial random variable and distribution are based on an experiment satisfying the following:

- 1) The experiment consists of a sequence of **trials**
- 2) Each trial can result in one of the same two possible outcomes, which we denote by **success (S)** or **failure (F)**.
- 3) The trials **are independent**.
- 4) The probability of success (**p**) is **constant** from trial to trial. So $P(S \text{ on trial } i) = p$.
- 5) The experiment continues (trials are performed) until a total of r successes has been observed, where r is a specified positive integer.

The **negative binomial random variable X** associated with this experiment is defined as

X = the number of F's that precede the r^{th} S.

Possible values of X are $0, 1, 2, \dots$. Let $nb(x; r, p)$ denote the pmf of X . The event $\{X = x\}$ is equivalent to $\{r - 1 \text{ S's in the first } (x + r - 1) \text{ trials and an S on the } (x + r)\text{th trial}\}$.

The pmf of the negative binomial rv X with parameters $r = \text{number of S's}$ and $p = P(S)$ is

$$nb(x; r, p) = \binom{x + r - 1}{r - 1} p^r (1 - p)^x \quad x = 0, 1, 2, \dots$$

For the same experiment, an alternative definition of the negative binomial random variable we see in the literature is as the following:

Y = the number of the trial on which the r^{th} S occurs

Since **Y = X + r**, the pmf of Y is given as:

$$p(y) = \binom{y - 1}{r - 1} p^r q^{y-r}, \quad y = r, r + 1, r + 2, \dots, \quad 0 \leq p \leq 1.$$

The mean, the variance and mgf of Y will be **different** than those of X :

$$\mu = E(Y) = \frac{r}{p} \quad \text{and} \quad \sigma^2 = V(Y) = \frac{r(1 - p)}{p^2}.$$

If $Y = X + r$, $E(Y) = E(X) + r = \frac{r(1-p)}{p} + r = \frac{r}{p}$
and $m_Y(t) = e^{rt} m_X(t)$.

The Hypergeometric Probability Distribution

Assumptions leading to the hypergeometric distribution:

1. The population or set to be sampled consists of N individuals, objects, or elements (a *finite* population).
2. Each individual can be characterized as a success (S) or a failure (F), and there are M successes in the population.
3. A sample of n individuals is selected **without replacement** in such a way that each subset of size n is equally likely to be chosen.

The random variable of interest is **$X = \text{the number of } S\text{'s in the sample}$** . The probability distribution of X depends on the parameters n , M , and N , so we wish to obtain $P(X = x) = h(x; n, M, N)$.

If X is the number of S 's in a completely random sample of size n drawn from a population consisting of M S 's and $(N - M)$ F 's, then the probability distribution of X , called the **hypergeometric distribution**, is given by

$$P(X = x) = h(x; n, M, N) = \frac{\binom{M}{x} \binom{N-M}{n-x}}{\binom{N}{n}}$$

for x an integer satisfying $\max(0, n - N + M) \leq x \leq \min(n, M)$.

Recall the Binomial experiment:

- 1) The experiment consists of a sequence of n smaller experiments called **trials**, where n is fixed in advance of the experiment.
- 2) Each trial can result in one of the same two possible outcomes, which we denote by **success** (S) or **failure** (F).
- 3) The trials **are independent**.
- 4) The probability of success (p) is **constant** from trial to trial.

The **binomial random variable X** associated with this experiment is defined as
 $X = \text{the number of } S\text{'s among the } n \text{ trials.}$

$$b(x; n, p) = \begin{cases} \binom{n}{x} p^x (1-p)^{n-x} & x = 0, 1, 2, \dots, n \\ 0 & \text{otherwise} \end{cases}$$

The Hypergeometric and Binomial Probability Distributions

In general, if the sample size n is smaller than the number of successes in the population (M), then the largest possible X value is n . However, if $M < n$ (e.g., a sample size of 25 and only 15 successes in the population), then X can be at most M . Similarly, whenever the number of population failures ($N - M$) exceeds the sample size, the smallest possible X value is 0 (since all sampled individuals might then be failures). However, if $N - M < n$, the smallest possible X value is $n - (N - M)$. Summarizing, the possible values of the hypergeometric rv X satisfy the restriction $\max[0, n - (N - M)] \leq x \leq \min(n, M)$.

The Mean and Variance of X :

Proposition:

The mean and variance of the hypergeometric rv X having pmf $h(x, n, M, N)$ are

$$E(X) = n \cdot \frac{M}{N} \quad V(X) = \left(\frac{N - n}{N - 1} \right) \cdot n \cdot \frac{M}{N} \left(1 - \frac{M}{N} \right)$$

The ratio M/N is the proportion of S 's in the population. Replacing M/N by p in $E(X)$ and $V(X)$ gives

$$E(X) = np$$
$$V(X) = \left(\frac{N - n}{N - 1} \right) \cdot np(1 - p) \quad (3.16)$$

Expression (3.16) shows that the means of the binomial and hypergeometric rv's are equal, whereas the variances of the two rv's differ by the factor $(N - n)/(N - 1)$, often called the **finite population correction factor**. This factor is < 1 , so the hypergeometric variable has smaller variance than does the binomial rv. The correction factor can be written $(1 - n/N)/(1 - 1/N)$, which is approximately 1 when n is small relative to N .

Our general rule of thumb stated that if sampling is without replacement but n/N is at most .05, then the binomial distribution can be used to compute approximate probabilities involving the number of S 's in the sample. A more precise statement is as follows: Let the population size, N , and number of population S 's, M , get large with the ratio M/N approaching p . Then $h(x; n, M, N)$ approaches $b(x; n, p)$; so for n/N small, the two are approximately equal provided that p is not too near either 0 or 1. This is the rationale for our rule of thumb.

The Poisson Probability Distribution

The Poisson probability distribution often provides a good model for the probability distribution of the number Y of rare events that occur in space, time, volume, or any other dimension, where λ is the average value of Y .

Definition:

A random variable X is said to have a **Poisson distribution** with parameter λ ($\lambda > 0$) if the pmf of X is

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

Proposition:

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

Proof Begin with the binomial pmf:

$$\begin{aligned} b(x; n, p) &= \binom{n}{x} p^x (1-p)^{n-x} = \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x} \\ &= \frac{n \cdot (n-1) \cdot \dots \cdot (n-x+1)}{x!} p^x (1-p)^{n-x} \end{aligned}$$

Include n^x in both the numerator and denominator:

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

Taking the limit as $n \rightarrow \infty$ and $p \rightarrow 0$ with $np \rightarrow \lambda$,

$$\lim_{n \rightarrow \infty} b(x; n, p) = 1 \cdot 1 \cdot \dots \cdot 1 \cdot \frac{\lambda^x}{x!} \cdot \left(\lim_{n \rightarrow \infty} \frac{(1 - np/n)^n}{1} \right)$$

The limit on the right can be obtained from the calculus theorem that says the limit of $(1 - a_n/n)^n$ is e^{-a} if $a_n \rightarrow a$. Because $np \rightarrow \lambda$,

$$\lim_{n \rightarrow \infty} b(x; n, p) = \frac{\lambda^x}{x!} \cdot \lim_{n \rightarrow \infty} \left(1 - \frac{np}{n} \right)^n = \frac{\lambda^x e^{-\lambda}}{x!} = p(x; \lambda) \quad \blacksquare$$

According to the proposition, in any binomial experiment for which n is large and p is small, $b(x; n, p) \approx p(x; \lambda)$ where $\lambda = np$. As a rule of thumb, this approximation can safely be applied if $n > 50$ and $np < 5$.

Proposition:

Since $b(x; n, p) \rightarrow p(x; \lambda)$ as $n \rightarrow \infty$, $p \rightarrow 0$, $np \rightarrow \lambda$, the mean and variance of a binomial variable should approach those of a Poisson variable. These limits are $np \rightarrow \lambda$ and $np(1-p) \rightarrow \lambda$.

If X has a Poisson distribution with parameter λ , then $E(X) = V(X) = \lambda$.

These results can also be derived directly from the definitions of mean and variance.

Proposition:

The Poisson moment generating function is

$$M_X(t) = e^{\lambda(e^t - 1)}$$

Proof The mgf is by definition

$$M_X(t) = E(e^{tx}) = \sum_{x=0}^{\infty} e^{tx} e^{-\lambda} \frac{\lambda^x}{x!} = e^{-\lambda} \sum_{x=0}^{\infty} \frac{(\lambda e^t)^x}{x!} = e^{-\lambda} e^{\lambda e^t} = e^{\lambda e^t - \lambda}$$

This uses the series expansion $\sum_{x=0}^{\infty} u^x / x! = e^u$. ■

EXAMPLES

Example 1 : Binomial Probability Distribution

Example: Experience has shown that 30% of all persons afflicted by a certain illness recover. A drug company has developed a new medication. Ten people with the illness were selected at random and received the medication; nine recovered shortly thereafter. Suppose that the medication was absolutely worthless. What is the probability that at least nine of ten receiving the medication will recover?

Solution: Let Y denote the number of people who recover. If the medication is worthless, the probability that a single ill person will recover is $p = .3$. Then the number of trials is $n = 10$ and the probability of *exactly* nine recoveries is

$$P(Y = 9) = p(9) = \binom{10}{9} (.3)^9 (.7) = .000138.$$

Similarly, the probability of exactly ten recoveries is

$$P(Y = 10) = p(10) = \binom{10}{10} (.3)^{10} (.7)^0 = .000006,$$

and

$$P(Y \geq 9) = p(9) + p(10) = .000138 + .000006 = .000144.$$

If the medication is ineffective, the probability of observing at least nine recoveries is extremely small. If we administered the medication to ten individuals and observed at least nine recoveries, then either (1) the medication is worthless and we have observed a rare event or (2) the medication is indeed useful in curing the illness. We adhere to conclusion 2. ■

Example 2 : Binomial Probability Distribution

Example: Suppose that a lot of 5000 electrical fuses contains 5% defectives. If a sample of 5 fuses is tested, find the probability of observing at least one defective.

If $n = 20$ fuses are randomly sampled from this lot, find the probability that at least four defectives will be observed.

Solution: It is reasonable to assume that Y , the number of defectives observed, has an approximate binomial distribution because the lot is large. Removing a few fuses does not change the composition of those remaining enough to cause us concern. Thus,

$$\begin{aligned} P(\text{at least one defective}) &= 1 - p(0) = 1 - \binom{5}{0} p^0 q^5 \\ &= 1 - (.95)^5 = 1 - .774 = .226. \end{aligned}$$

Notice that there is a fairly large chance of seeing at least one defective, even though the sample is quite small.

Letting Y denote the number of defectives in the sample, we assume the binomial model for Y , with $p = .05$. Thus,

$$P(Y \geq 4) = 1 - P(Y \leq 3),$$

and using Table 1, Appendix 3 [or the *R* command `pbinom(3, 20, .05)`], we obtain

$$P(Y \leq 3) = \sum_{y=0}^3 p(y) = .984.$$

The value .984 is found in the table labeled $n = 20$ in Table 1, Appendix 3. Specifically, it appears in the column labeled $p = .05$ and in the row labeled $a = 3$. It follows that

$$P(Y \geq 4) = 1 - .984 = .016.$$

This probability is quite small. If we did indeed observe more than three defectives out of 20 fuses, we might suspect that the reported 5% defect rate is erroneous. ■

Table 1 (Continued)

(c) $n = 15$

a	p													a
	0.01	0.05	0.10	0.20	0.30	0.40	0.50	0.60	0.70	0.80	0.90	0.95	0.99	
0	.860	.463	.206	.035	.005	.000	.000	.000	.000	.000	.000	.000	.000	0
1	.990	.829	.549	.167	.035	.005	.000	.000	.000	.000	.000	.000	.000	1
2	1.000	.964	.816	.398	.127	.027	.004	.000	.000	.000	.000	.000	.000	2
3	1.000	.995	.944	.648	.297	.091	.018	.002	.000	.000	.000	.000	.000	3
4	1.000	.999	.987	.836	.515	.217	.059	.009	.001	.000	.000	.000	.000	4
5	1.000	1.000	.998	.939	.722	.403	.151	.034	.004	.000	.000	.000	.000	5
6	1.000	1.000	1.000	.982	.869	.610	.304	.095	.015	.001	.000	.000	.000	6
7	1.000	1.000	1.000	.996	.950	.787	.500	.213	.050	.004	.000	.000	.000	7
8	1.000	1.000	1.000	.999	.985	.905	.696	.390	.131	.018	.000	.000	.000	8
9	1.000	1.000	1.000	1.000	.996	.966	.849	.597	.278	.061	.002	.000	.000	9
10	1.000	1.000	1.000	1.000	.999	.991	.941	.783	.485	.164	.013	.001	.000	10
11	1.000	1.000	1.000	1.000	1.000	.998	.982	.909	.703	.352	.056	.005	.000	11
12	1.000	1.000	1.000	1.000	1.000	1.000	.996	.973	.873	.602	.184	.036	.000	12
13	1.000	1.000	1.000	1.000	1.000	1.000	1.000	.995	.965	.833	.451	.171	.010	13
14	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	.995	.965	.794	.537	.140	14

(d) $n = 20$

a	p													a
	0.01	0.05	0.10	0.20	0.30	0.40	0.50	0.60	0.70	0.80	0.90	0.95	0.99	
0	.818	.358	.122	.012	.001	.000	.000	.000	.000	.000	.000	.000	.000	0
1	.983	.736	.392	.069	.008	.001	.000	.000	.000	.000	.000	.000	.000	1
2	.999	.925	.677	.206	.035	.004	.000	.000	.000	.000	.000	.000	.000	2
3	1.000	.984	.867	.411	.107	.016	.001	.000	.000	.000	.000	.000	.000	3
4	1.000	.997	.957	.630	.238	.051	.006	.000	.000	.000	.000	.000	.000	4
5	1.000	1.000	.989	.804	.416	.126	.021	.002	.000	.000	.000	.000	.000	5
6	1.000	1.000	.998	.913	.608	.250	.058	.006	.000	.000	.000	.000	.000	6
7	1.000	1.000	1.000	.968	.772	.416	.132	.021	.001	.000	.000	.000	.000	7
8	1.000	1.000	1.000	.990	.887	.596	.252	.057	.005	.000	.000	.000	.000	8
9	1.000	1.000	1.000	.997	.952	.755	.412	.128	.017	.001	.000	.000	.000	9
10	1.000	1.000	1.000	.999	.983	.872	.588	.245	.048	.003	.000	.000	.000	10
11	1.000	1.000	1.000	1.000	.995	.943	.748	.404	.113	.010	.000	.000	.000	11
12	1.000	1.000	1.000	1.000	.999	.979	.868	.584	.228	.032	.000	.000	.000	12
13	1.000	1.000	1.000	1.000	1.000	.994	.942	.750	.392	.087	.002	.000	.000	13
14	1.000	1.000	1.000	1.000	1.000	.998	.979	.874	.584	.196	.011	.000	.000	14
15	1.000	1.000	1.000	1.000	1.000	1.000	.994	.949	.762	.370	.043	.003	.000	15
16	1.000	1.000	1.000	1.000	1.000	1.000	.999	.984	.893	.589	.133	.016	.000	16
17	1.000	1.000	1.000	1.000	1.000	1.000	1.000	.996	.965	.794	.323	.075	.001	17
18	1.000	1.000	1.000	1.000	1.000	1.000	1.000	.999	.992	.931	.608	.264	.017	18
19	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	.999	.988	.878	.642	.182	19

Example 3: Binomial Probability Distribution

Prove the following using the definition of expected value:

Let Y be a binomial random variable based on n trials and success probability p . Then

$$\mu = E(Y) = np \quad \text{and} \quad \sigma^2 = V(Y) = npq.$$

By Definitions

$$E(Y) = \sum_y yp(y) = \sum_{y=0}^n y \binom{n}{y} p^y q^{n-y}.$$

Notice that the first term in the sum is 0 and hence that

$$\begin{aligned} E(Y) &= \sum_{y=1}^n y \frac{n!}{(n-y)!y!} p^y q^{n-y} \\ &= \sum_{y=1}^n \frac{n!}{(n-y)!(y-1)!} p^y q^{n-y}. \end{aligned}$$

The summands in this last expression bear a striking resemblance to binomial probabilities. In fact, if we factor np out of each term in the sum and let $z = y - 1$,

$$\begin{aligned} E(Y) &= np \sum_{y=1}^n \frac{(n-1)!}{(n-y)!(y-1)!} p^{y-1} q^{n-y} \\ &= np \sum_{z=0}^{n-1} \frac{(n-1)!}{(n-1-z)!z!} p^z q^{n-1-z} \\ &= np \sum_{z=0}^{n-1} \binom{n-1}{z} p^z q^{n-1-z}. \end{aligned}$$

Notice that $p(z) = \binom{n-1}{z} p^z q^{n-1-z}$ is the binomial probability function based on $(n-1)$ trials. Thus, $\sum_z p(z) = 1$, and it follows that

$$\mu = E(Y) = np.$$

we know that $\sigma^2 = V(Y) = E(Y^2) - \mu^2$. Thus, σ^2 can be calculated if we find $E(Y^2)$. Finding $E(Y^2)$ directly is difficult because

$$E(Y^2) = \sum_{y=0}^n y^2 p(y) = \sum_{y=0}^n y^2 \binom{n}{y} p^y q^{n-y} = \sum_{y=0}^n y^2 \frac{n!}{y!(n-y)!} p^y q^{n-y}$$

and the quantity y^2 does not appear as a factor of $y!$. Where do we go from here? Notice that

$$E[Y(Y-1)] = E(Y^2 - Y) = E(Y^2) - E(Y)$$

and, therefore,

$$E(Y^2) = E[Y(Y-1)] + E(Y) = E[Y(Y-1)] + \mu.$$

In this case,

$$E[Y(Y-1)] = \sum_{y=0}^n y(y-1) \frac{n!}{y!(n-y)!} p^y q^{n-y}.$$

The first and second terms of this sum equal zero (when $y = 0$ and $y = 1$). Then

$$E[Y(Y-1)] = \sum_{y=2}^n \frac{n!}{(y-2)!(n-y)!} p^y q^{n-y}.$$

(Notice the cancellation that led to this last result. The anticipation of this cancellation is what actually motivated the consideration of $E[Y(Y-1)]$.) Again, the summands in the last expression look very much like binomial probabilities. Factor $n(n-1)p^2$ out of each term in the sum and let $z = y - 2$ to obtain

$$\begin{aligned} E[Y(Y-1)] &= n(n-1)p^2 \sum_{y=2}^n \frac{(n-2)!}{(y-2)!(n-y)!} p^{y-2} q^{n-y} \\ &= n(n-1)p^2 \sum_{z=0}^{n-2} \frac{(n-2)!}{z!(n-2-z)!} p^z q^{n-2-z} \\ &= n(n-1)p^2 \sum_{z=0}^{n-2} \binom{n-2}{z} p^z q^{n-2-z}. \end{aligned}$$

Again note that $p(z) = \binom{n-2}{z} p^z q^{n-2-z}$ is the binomial probability function based on $(n-2)$ trials. Then $\sum_{z=0}^{n-2} p(z) = 1$ (again using the device illustrated in the derivation of the mean) and

$$E[Y(Y-1)] = n(n-1)p^2.$$

Thus,

$$E(Y^2) = E[Y(Y-1)] + \mu = n(n-1)p^2 + np$$

and

$$\begin{aligned} \sigma^2 &= E(Y^2) - \mu^2 = n(n-1)p^2 + np - n^2p^2 \\ &= np[(n-1)p + 1 - np] = np(1-p) = npq. \end{aligned}$$

Examples 4-5-6: Geometric and Negative Binomial Probability Distributions

Example 4: Geometric

Consider the experiment of tossing a fair die until a 2 is observed. What is the probability that this experiment ends at the 5th toss?

Solution: If A is the event of observing a 2 on a toss, the outcome should be B= A'A'A'A. We have

$$p(B) = \left(\frac{5}{6}\right)^4 \left(\frac{1}{6}\right)$$

Example 5: Negative Binomial

A geological study indicates that an exploratory oil well drilled in a particular region should strike oil with probability .2. Find the probability that the third oil strike comes on the fifth well drilled.

Solution: Assuming independent drillings and probability .2 of striking oil with any one well, let Y denote the number of the trial on which the third oil strike occurs. Then it is reasonable to assume that Y has a negative binomial distribution with $p = .2$. Because we are interested in $r = 3$ and $y = 5$,

$$\begin{aligned} P(Y = 5) &= p(5) = \binom{4}{2} (.2)^3 (.8)^2 \\ &= 6(.008)(.64) = .0307. \end{aligned}$$

Example 6: Negative Binomial

A pediatrician wishes to recruit 5 couples, each of whom is expecting their first child, to participate in a new natural childbirth regimen. Let $p = P(\text{a randomly selected couple agrees to participate})$. If $p = .2$, what is the probability that 15 couples must be asked before 5 are found who agree to participate? That is, with $S = \{\text{agrees to participate}\}$, what is the probability that 10 F 's occur before the fifth S ? Substituting $r = 5$, $p = .2$, and $x = 10$ into $nb(x; r, p)$ gives

$$nb(10; 5, 2) = \binom{14}{4} .2^5 .8^{10} = .034$$

The probability that at most 10 F 's are observed (at most 15 couples are asked) is

$$P(X \leq 10) = \sum_{x=0}^{10} nb(x; 5, 2) = .2^5 \sum_{x=0}^{10} \binom{x+4}{4} .8^x = .164$$

Example 7: Hypergeometric Probability Distribution

An industrial product is shipped in lots of 20. Testing to determine whether an item is defective is costly, and hence the manufacturer samples his production rather than using a 100% inspection plan. A sampling plan, constructed to minimize the number of defectives shipped to customers, calls for sampling five items from each lot and rejecting the lot if more than one defective is observed. (If the lot is rejected, each item in it is later tested.) If a lot contains four defectives, what is the probability that it will be rejected? What is the expected number of defectives in the sample of size 5? What is the variance of the number of defectives in the sample of size 5?

Solution: (Note that $\frac{n}{N} = \frac{5}{20}$ which is greater than 0.05 so binomial distribution can't be used. In the notation used below $M=r$.)

Let Y equal the number of defectives in the sample. Then $N = 20$, $r = 4$, and $n = 5$. The lot will be rejected if $Y = 2, 3$, or 4. Then

$$\begin{aligned}P(\text{rejecting the lot}) &= P(Y \geq 2) = p(2) + p(3) + p(4) \\&= 1 - p(0) - p(1) \\&= 1 - \frac{\binom{4}{0}\binom{16}{5}}{\binom{20}{5}} - \frac{\binom{4}{1}\binom{16}{4}}{\binom{20}{5}} \\&= 1 - .2817 - .4696 = .2487.\end{aligned}$$

The mean and variance of the number of defectives in the sample of size 5 are

$$\mu = \frac{(5)(4)}{20} = 1 \quad \text{and} \quad \sigma^2 = 5 \left(\frac{4}{20} \right) \left(\frac{20-4}{20} \right) \left(\frac{20-5}{20-1} \right) = .632. \quad \blacksquare$$

Examples 8-9 : Poisson and Binomial Probability Distributions

Example 8 : Poisson and Binomial

If a publisher of nontechnical books takes great pains to ensure that its books are free of typographical errors, so that the probability of any given page containing at least one such error is .005 and errors are independent from page to page, what is the probability that one of its 400-page novels will contain exactly one page with errors? At most three pages with errors?

With S denoting a page containing at least one error and F an error-free page, the number X of pages containing at least one error is a binomial rv with $n = 400$ and $p = .005$, so $np = 2$. We wish

$$P(X = 1) = b(1; 400, .005) \approx p(1; 2) = \frac{e^{-2}2^1}{1!} = .270671$$

The binomial value is $b(1; 400, .005) = .270669$, so the approximation is good to five decimal places here.

Similarly,

$$\begin{aligned} P(X \leq 3) &\approx \sum_{x=0}^3 p(x; 2) = \sum_{x=0}^3 e^{-2} \frac{2^x}{x!} \\ &= .135335 + .270671 + .270671 + .180447 \\ &= .8571 \end{aligned}$$

and this again is quite close to the binomial value $P(X \leq 3) = .8576$. ■

Example 9 : Poisson

Prove the following using the definition of expected value:

If Y is a random variable possessing a Poisson distribution with parameter λ , then

$$\mu = E(Y) = \lambda$$

Proof: By definition,

$$E(Y) = \sum_y y p(y) = \sum_{y=0}^{\infty} y \frac{\lambda^y e^{-\lambda}}{y!}.$$

Notice that the first term in this sum is equal to 0 (when $y = 0$), and, hence,

$$E(Y) = \sum_{y=1}^{\infty} y \frac{\lambda^y e^{-\lambda}}{y!} = \sum_{y=1}^{\infty} \frac{\lambda^y e^{-\lambda}}{(y-1)!}.$$

As it stands, this quantity is not equal to the sum of the values of a probability function $p(y)$ over all values of y , but we can change it to the proper form by factoring λ out of the expression and letting $z = y - 1$. Then the limits of summation become $z = 0$ (when $y = 1$) and $z = \infty$ (when $y = \infty$), and

$$E(Y) = \lambda \sum_{y=1}^{\infty} \frac{\lambda^{y-1} e^{-\lambda}}{(y-1)!} = \lambda \sum_{z=0}^{\infty} \frac{\lambda^z e^{-\lambda}}{z!}.$$

Notice that $p(z) = \lambda^z e^{-\lambda} / z!$ is the probability function for a Poisson random variable, and $\sum_{z=0}^{\infty} p(z) = 1$. Therefore, $E(Y) = \lambda$. Thus, the mean of a Poisson random variable is the single parameter λ that appears in the expression for the Poisson probability function.

Examples 10-11 : Poisson Probability Distribution

Example 10: A certain type of tree has seedlings randomly dispersed in a large area, with the mean density of seedlings being approximately five per square yard. If a forester randomly locates ten 1-square-yard sampling regions in the area, find the probability that none of the regions will contain seedlings.

Solution: If the seedlings really are randomly dispersed, the number of seedlings per region, Y , can be modeled as a Poisson random variable with $\lambda = 5$. (The average density is five per square yard.) Thus,

$$P(Y = 0) = p(0) = \frac{\lambda^0 e^{-\lambda}}{0!} = e^{-5} = .006738.$$

The probability that $Y = 0$ on ten independently selected regions is $(e^{-5})^{10}$ because the probability of the intersection of independent events is equal to the product of the respective probabilities. The resulting probability is extremely small. Thus, if this event actually occurred, we would seriously question the assumption of randomness, the stated average density of seedlings, or both. ■

Example 11: Suppose that Y is a random variable with moment-generating function $m_Y(t) = e^{3.2(e^t - 1)}$. What is the distribution of Y ?

Solution: we showed that the moment-generating function of a Poisson distributed random variable with mean λ is $m(t) = e^{\lambda(e^t - 1)}$. Note that the moment-generating function of Y is exactly equal to the moment-generating function of a Poisson distributed random variable with $\lambda = 3.2$. Because moment-generating functions are unique, Y *must* have a Poisson distribution with mean 3.2. ■

References

- Devore, J.L. and Berk, K.N. (2012) Modern Mathematical Statistics with Applications, 2nd Edition, New York: Springer.
- Wackerly, D.D., Mendenhall, W. and Scheaffer, R.L., (2008) Mathematical Statistics With Applications, 7th Edition, Pacific Grove: Duxbury.