Mathematical Statistics and Data Analysis: Lection 5.Distributions and their characteristics

Aleksei Zhukov

Data Scientist, Research Fellow

BI BRICS,
2020

Outline

The **Poisson frequency function** with parameter λ ($\lambda > 0$) is

$$P(X = k) = \frac{\lambda^k}{k!} e^{-\lambda}, \qquad k = 0, 1, 2, \ldots$$

The **Poisson distribution** can be derived as the limit of a binomial distribution as the number of trials, n, approaches infinity and the probability of success on each trial, p, approaches zero in such a way that $np = \lambda$. The binomial frequency function is

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

Setting $np = \lambda$, this expression becomes

$$\begin{aligned} p(k) &= \frac{n!}{k!(n-k)!} \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{n-k} \\ &= \frac{\lambda^k}{k!} \frac{n!}{(n-k)!} \frac{1}{n^k} \left(1 - \frac{\lambda}{n}\right)^n \left(1 - \frac{\lambda}{n}\right)^{-k} \end{aligned}$$

As $n \to \infty$,

$$\frac{\lambda}{n} \to 0$$

$$\frac{n!}{(n-k)!n^k} \to 1$$

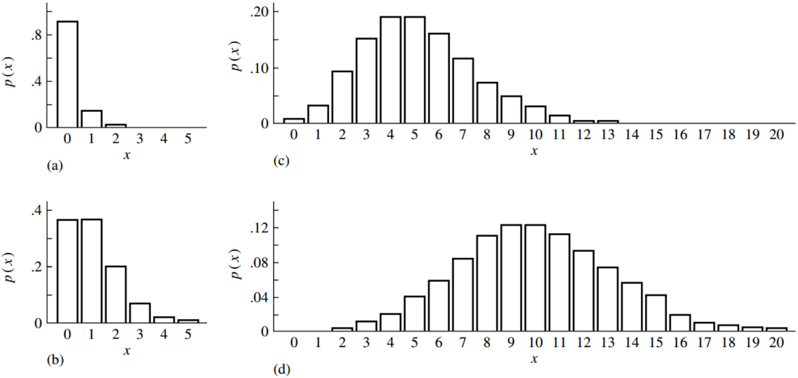
$$\left(1 - \frac{\lambda}{n}\right)^n \to e^{-\lambda}$$

$$\left(1 - \frac{\lambda}{n}\right)^{-k} \to 1$$

=>

We have the Poisson frequency function.

$$p(k) \to \frac{\lambda^k e^{-\lambda}}{k!}$$



Poisson frequency functions, (a) $\lambda = .1$, (b) $\lambda = 1$, (c) $\lambda = 5$, (d) $\lambda = 10$.

Two dice are rolled 100 times, and the number of double sixes, X, is counted. The distribution of X is binomial with $n = 100$ and $p = \frac{1}{36} = .0278$. Since n is large and p is small, we can approximate the binomial probabilities by Poisson probabilities with $\lambda = np = 2.78$. The exact binomial probabilities and the Poisson approximations are shown in the following table:

k	Binomial Probability	Poisson Approximation
0	.0596	.0620
1	.1705	.1725
2	.2414	.2397
3	.2255	.2221
4	.1564	.1544
5	.0858	.0858
6	.0389	.0398
7	.0149	.0158
8	.0050	.0055
9	.0015	.0017
10	.0004	.0005
11	.0001	.0001

Example

The Poisson distribution has been used as a model by insurance companies. For example, the number of freak accidents, such as falls in the shower, for a large population of people in a given time period might be modeled as a Poisson distribution, because the accidents would presumably be rare and independent (provided there was only one person in the shower :)).

In applications, we are often interested in random variables that can take on a continuum of values rather than a finite or countably infinite number.

For a continuous random variable, the role of the frequency function is taken by a **density function,** $f(x)$, which has the properties that $f(x) \geq 0$, f is piecewise continuous, and $\int_{-\infty}^{\infty} f(x)\, dx = 1$. If X is a random variable with a density function f, then for any $a < b$, the probability that X falls in the interval (a, b) is the area under the density function between a and b:

$$P(a < X < b) = \int_a^b f(x)\, dx$$

The cumulative distribution function of a continuous random variable X is defined in the same way as for a discrete random variable:

$$F(x) = P(X \leq x)$$

$F(x)$ can be expressed in terms of the density function:

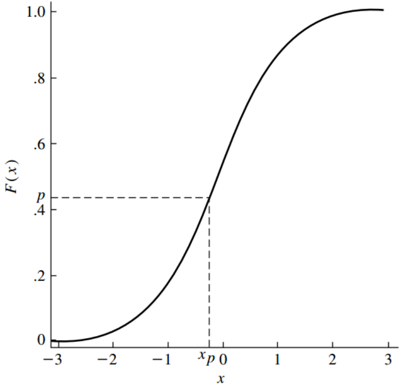
$$F(x) = \int_{-\infty}^{x} f(u)\, du$$

From the fundamental theorem of calculus, if f is continuous at x, $f(x) = F'(x)$.

The cdf can be used to evaluate the probability that X falls in an interval:

$$P(a \leq X \leq b) = \int_{a}^{b} f(x)\, dx = F(b) - F(a)$$

Suppose that F is the cdf of a continuous random variable and is strictly increasing on some interval I, and that $F = 0$ to the left of I and $F = 1$ to the right of I; I may be unbounded. Under this assumption, the inverse function F^{-1} is well defined; $x = F^{-1}(y)$ if $y = F(x)$. The pth **quantile** of the distribution F is defined to be that value x_p such that $F(x_p) = p$, or $P(X \leq x_p) = p$. Under the preceding assumption stated, x_p is uniquely defined as $x_p = F^{-1}(p)$; see Figure. Special cases are $p = \frac{1}{2}$, which corresponds to the **median** of F; and $p = \frac{1}{4}$ and $p = \frac{3}{4}$, which correspond to the lower and upper **quartiles** of F.



Value at Risk

Financial firms need to quantify and monitor the risk of their investments. **Value at Risk (VaR)** is a widely used measure of potential losses. It involves two parameters: a time horizon and a level of confidence. For example, if the VaR of an institution is \$10 million with a one-day horizon and a level of confidence of 95%, the interpretation is that there is a 5% chance of losses exceeding \$10 million. Such a loss should be anticipated about once in 20 days.

To see how VaR is computed, suppose the current value of the investment is V_0 and the future value is V_1. The return on the investment is $R = (V_1 - V_0)/V_0$, which is modeled as a continuous random variable with cdf $F_R(r)$. Let the desired level of confidence be denoted by $1 - \alpha$. We want to find v^*, the VaR. Then

$$\begin{aligned}\alpha &= P(V_0 - V_1 \geq v^*)\\ &= P\left(\frac{V_1 - V_0}{V_0} \leq -\frac{v^*}{V_0}\right)\\ &= F_R\left(-\frac{v^*}{V_0}\right)\end{aligned}$$

Thus, $-v^*/V_0$ is the α quantile, r_α; and $v^* = -V_0 r_\alpha$. The VaR is minus the current value times the α quantile of the return distribution. ■

A **uniform random variable** on the interval $[0, 1]$ is a model for what we mean when we say "choose a number at random between 0 and 1." Any real number in the interval is a possible outcome, and the probability model should have the property that the probability that X is in any subinterval of length h is equal to h. The following density function does the job:

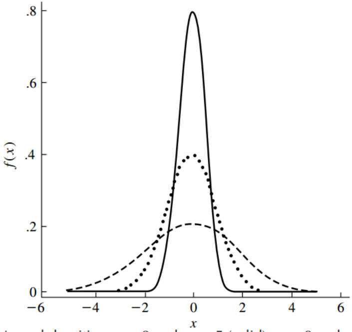
$$f(x) = \begin{cases} 1, & 0 \leq x \leq 1 \\ 0, & x < 0 \text{ or } x > 1 \end{cases}$$

This is called the **uniform density** on $[0, 1]$. The uniform density on a general interval $[a, b]$ is

$$f(x) = \begin{cases} 1/(b-a), & a \leq x \leq b \\ 0, & x < a \text{ or } x > b \end{cases}$$

The normal distribution plays a central role in probability and statistics. This distribution is also called the Gaussian distribution after Carl Friedrich Gauss, who proposed it as a model for measurement errors. The normal distribution has been used as a model for such diverse phenomena as a person's height, the distribution of IQ scores, and so on. The density function of the normal distribution depends on two parameters, μ and σ (where $-\infty < \mu < \infty, \sigma > 0$):

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2}$$

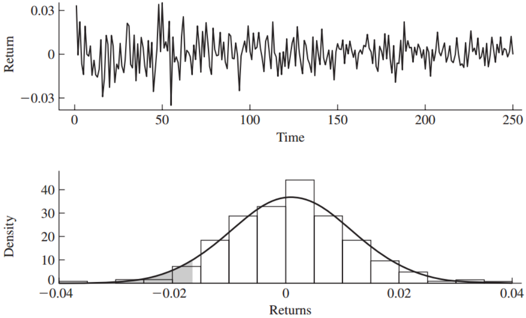


Normal densities, $\mu = 0$ and $\sigma = .5$ (solid), $\mu = 0$ and $\sigma = 1$ (dotted), and $\mu = 0$ and $\sigma = 2$ (dashed).

The cdf cannot be evaluated in closed form from this density function (the integral that defines the cdf cannot be evaluated by an explicit formula but must be found numerically).
As shorthand for the statement "X follows a normal distribution with parameters μ and σ", it is convenient to use $X \sim N(\mu, \sigma^2)$. From the form of the density function, we see that the density is symmetric about μ, $f(\mu - x) = f(\mu + x)$, where it has a maximum, and that the rate at which it falls off is determined by σ . Figure on the previous slide shows several normal densities. The special case for which $\mu = 0$ and $\sigma = 1$ is called the standard normal density. Its cdf is denoted by Φ and its density by ϕ.

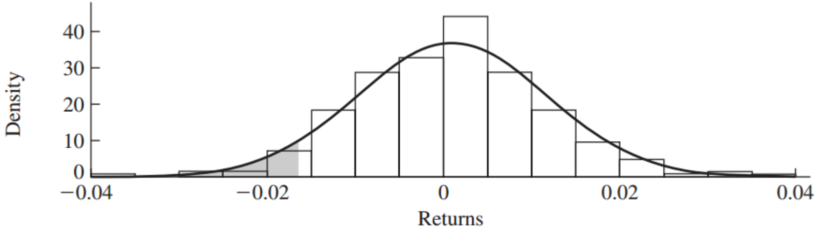
S&P 500 Example

The Standard and Poors 500 is an index of important U.S. stocks; each stock's weight in the index is proportional to its market value. Individuals can invest in mutual funds that track the index. The top panel of figure shows the sequential values of the returns during 2003. The average return during this period was 0.1% per day, and we can see from the figure that daily fluctuations were as large as 3% or 4%. The lower panel of the figure shows a histogram of the returns and a fitted normal density with $\mu = 0.001$ and $\sigma = 0.01$.



S&P 500 Example

A financial company could use the fitted normal density in calculating its Value at Risk. Using a time horizon of one day and a confidence level of 95%, the VaR is the current investment in the index, V_0, multiplied by the negative of the 0.05 quantile of the distribution of returns. In this case, the quantile can be calculated to be -0.0165, so the VaR is $0.0165V_0$. Thus, if V_0 is \$10 million, the VaR is \$165,000. The company can have 95% "confidence" that its losses will not exceed that amount on a given day. However, it should not be surprised if that amount is exceeded about once in every 20 trading days.



The concept of the expected value of a random variable parallels the notion of a weighted average. The possible values of the random variable are weighted by their probabilities, as specified in the following definition. $E(X)$ is also referred to as the mean of X and is often denoted by μ.

DEFINITION

If X is a discrete random variable with frequency function $p(x)$, the expected value of X, denoted by $E(X)$, is

$$E(X) = \sum_i x_i p(x_i)$$

provided that $\sum_i |x_i| p(x_i) < \infty$. If the sum diverges, the expectation is undefined. ■

A roulette wheel has the numbers 1 through 36, as well as 0 and 00. If you bet \$1 that an odd number comes up, you win or lose \$1 according to whether that event occurs. If X denotes your net gain, $X = 1$ with probability $\frac{18}{38}$ and $X = -1$ with probability $\frac{20}{38}$. The expected value of X is

$$E(X) = 1 \times \frac{18}{38} + (-1) \times \frac{20}{38} = -\frac{1}{19}$$

Thus, your expected loss is about \$.05.

St. Petersburg Paradox

A gambler has the following strategy for playing a sequence of games: He starts off betting \$1; if he loses, he doubles his bet; and he continues to double his bet until he finally wins. To analyze this scheme, suppose that the game is fair and that he wins or loses the amount he bets. At trial 0, he bets \$1; if he loses, he bets \$2 at trial 1; and if he has not won by the kth trial, he bets $\$2^k$. When he finally wins, he will be \$1 ahead, which can be checked by going through the scheme for the first few values of k. This seems like a foolproof way to win \$1. What could be wrong with it?

Let X denote the amount of money bet on the very last game (the game he wins). Because the probability that k losses are followed by one win is $2^{-(k+1)}$,

$$P(X = 2^k) = \frac{1}{2^{k+1}}$$

and

$$E(X) = \sum_{n=0}^{\infty} nP(X = n)$$
$$= \sum_{k=0}^{\infty} 2^k \frac{1}{2^{k+1}} = \infty$$

Formally, $E(X)$ is not defined. Practically, the analysis shows a flaw in this scheme, which is that it does not take into account the enormous amount of capital required. ■

Expected utility hypothesis

Expected utility hypothesis–concerning people's preferences with regard to choices that have uncertain outcomes (probabilistic)–states that the subjective value associated with an individual's gamble is the statistical expectation of that individual's valuations of the outcomes of that gamble, where these valuations may differ from the dollar value of those outcomes. The introduction of St. Petersburg Paradox by Daniel Bernoulli in 1738 is considered the beginnings of the hypothesis. This hypothesis has proven useful to explain some popular choices that seem to contradict the expected value criterion!

The definition of expectation for a continuous random variable is a fairly obvious extension of the discrete case – summation is replaced by integration.

DEFINITION

If X is a continuous random variable with density $f(x)$, then

$$E(X) = \int_{-\infty}^{\infty} x f(x)\, dx$$

provided that $\int |x|\, f(x) dx < \infty$. If the integral diverges, the expectation is undefined. ■

Normal Distribution

From the definition of the expectation, we have

$$E(X) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} x e^{-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}}\, dx$$

Making the change of variables $z = x - \mu$ changes this equation to

$$E(X) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} z e^{-z^2/2\sigma^2} dz + \frac{\mu}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-z^2/2\sigma^2} dz$$

The first integral is 0 since the contributions from $z < 0$ cancel those from $z > 0$, and the second integral is μ because the normal density integrates to 1. Thus,

$$E(X) = \mu$$

The parameter μ of the normal density is the expectation, or mean value. We could have made the derivation much shorter by claiming that it was "obvious" that since the center of symmetry of the density is μ, the expectation must be μ. ■

Let's introduces the **standard deviation** of a random variable, which is an indication of how dispersed the probability distribution is about its center, of how spread out on the average are the values of the random variable about its expectation. We first define the **variance** of a random variable Var(X) and the standard deviation σ. There's the following dependence $Var(X) = \sigma^2$.

DEFINITION

If X is a random variable with expected value $E(X)$, the variance of X is

$$\mathrm{Var}(X) = E\{[X - E(X)]^2\}$$

provided that the expectation exists. The standard deviation of X is the square root of the variance. ■

If X is a discrete random variable with frequency function $p(x)$ and expected value $\mu = E(X)$, then according to the definition and Theorem A of Section 4.1.1,

$$\mathrm{Var}(X) = \sum_i (x_i - \mu)^2 p(x_i)$$

whereas if X is a continuous random variable with density function $f(x)$ and $E(X) = \mu$

$$\mathrm{Var}(X) = \int_{-\infty}^{\infty} (x - \mu)^2 f(x)\, dx$$

THEOREM A

If $\mathrm{Var}(X)$ exists and $Y = a + bX$, then $\mathrm{Var}(Y) = b^2\mathrm{Var}(X)$.

Proof

Since $E(Y) = a + bE(X)$,

$$\begin{aligned} E[(Y - E(Y))^2] &= E\{[a + bX - a - bE(X)]^2\} \\ &= E\{b^2[X - E(X)]^2\} \\ &= b^2 E\{[X - E(X)]^2\} \\ &= b^2\mathrm{Var}(X) \end{aligned}$$

■

Bernoulli Distribution

If X has a Bernoulli distribution—that is, X takes on values 0 and 1 with probability $1 - p$ and p, respectively—then we have seen (Example A of Section 4.1.2) that $E(X) = p$. By the definition of variance,

$$\begin{aligned}\mathrm{Var}(X) &= (0-p)^2 \times (1-p) + (1-p)^2 \times p \\ &= p^2 - p^3 + p - 2p^2 + p^3 \\ &= p(1-p)\end{aligned}$$

Note that the expression $p(1 - p)$ is a quadratic with a maximum at $p = \frac{1}{2}$. If p is 0 or 1, the variance is 0, which makes sense since the probability distribution is concentrated at a single point and the random variable is not variable at all. The distribution is most dispersed when $p = \frac{1}{2}$. ■

Normal Distribution
We have seen that $E(X) = \mu$. Then

$$\mathrm{Var}(X) = E[(X-\mu)^2] = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} (x-\mu)^2 e^{-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}}\, dx$$

Making the change of variables $z = (x-\mu)/\sigma$ changes the right-hand side to

$$\frac{\sigma^2}{\sqrt{2\pi}} \int_{-\infty}^{\infty} z^2 e^{-z^2/2}\, dz$$

Finally, making the change of variables $u = z^2/2$ reduces the integral to a gamma function, and we find that $\mathrm{Var}(X) = \sigma^2$. ■

Please, add to the table from previous lesson Expectation and Variance columns.

Todays topics

1. Discrete Random Variables
   - Poisson Distribution
2. Continuous Random Variables
   - Uniform distribution
   - Gaussian distribution
3. Random Variables Characteristics
   - Expected value
   - Variance
4. Conclusions and Homework

Questions

Time for your questions!

zhukovalex13@gmail.com