Notes 4: Continuous Probability Distributions

Julio Garín
Department of Economics

Statistics for Economics

Spring 2012

Introduction

- What is the motivation for studying variables with infinite number of possible values.
- Examples.
- What we did so far?
- No 'point' probability here.

Describing a Continuous Random Variable

- Probability Density Function (PDF).
  - The PDF of a random variable x is defined as $f(x)$.
- What would be the Cumulative Distribution Function?
  - Before we used summations, what can we use now?
- Suppose x is a "measure" (for instance, distance). We may be interested in something like

$$P(x \leq a) = F(a) = ?$$

- What if we consider all values?

$$P(x \leq \infty) = F(\infty) = ?$$

Cumulative Distribution Function

- Some formalities:

Let X denote any random variable. The cumulative distribution function of X, denoted by $F(x)$, is such that $F(x) = P(X \leq x)$ for $-\infty < x < \infty$.

Properties of a Distribution Function: If $F(x)$ is a distribution function, then

1. $F(-\infty) \equiv \lim_{x \to -\infty} F(x) = 0$
2. $F(\infty) \equiv \lim_{x \to \infty} F(x) = 1$
3. $F(x)$ is a nondecreasing function of x.

A random variable X is said to be *continuous* if $F(x)$ is continuous, for $-\infty < x < \infty$.

Some Useful Relationships

Let $F(x)$ CDF for a continuous random variable X. Then $f(x)$, given by

$$f(x) = \frac{dF(x)}{dx} = F'(x)$$

wherever the derivative exists, is called, the *probability density function* for the random variable X.

- As before, many times will be easier to work with complements.

Properties of a Density Function: If $f(x)$ is a density function for a continuous r.v., then

1. $f(x) \geq 0 \quad \forall x \quad -\infty < x < \infty$.
2. $\int_{-\infty}^{\infty} f(x)dx = 1$.

- What if we want to calculate the probability that X falls in a specific interval?

Exercises

$$F(x) = \begin{cases} 0 & \text{for} \quad x < 0 \\ x & \text{for} \quad 0 \leq x \leq 1 \\ 1 & \text{for} \quad x > 1 \end{cases}$$

1. Find the probability density function of X and graph it.

Given $f(x) = cx^2$, with the support $0 \leq x \leq 2$, and $f(x) = 0$ elsewhere.

1. Find the value of c for which $f(x)$ is a valid density function.
2. Find $P(1 \leq X \leq 2)$.
3. Find $P(1 < X < 2)$

Expected Value

- The idea is the same as before.

The expected value of a continuous random variable X is

$$\mathbb{E}(X) = \int_{-\infty}^{\infty} xf(x)dx$$

provided that the integral exist.

- What about the variance?
  - Just use $\mathbb{E}(X^2) - \mu^2$.
- Calculate the expected value of the previous example.

Uniform Distribution

Uniform Distribution

- Suppose that the average points per game scored by the ND basketball, (P), is uniform over $[50, 100]$.
  - What does it mean?
  - What is the probability that the team will score between 60 and 80 points?
  - How to obtain the PDF?

Uniform Distribution

If $\theta_1 < \theta_2$, a r.v. X is said to have a continuous *uniform probability distribution* on the interval (θ_1, θ_2) iff the density function of X is

$$f(x) = \begin{cases} \frac{1}{\theta_2 - \theta_1} & \text{for } \theta_1 \leq \theta_2 \\ 0 & \text{elsewhere} \end{cases}$$

- Graphically.

If $\theta_1 < \theta_2$, and X is r.v. uniformly distributed on the interval (θ_1, θ_2) then

$$\mu = \mathbb{E}(X) = \frac{\theta_2 + \theta_1}{2}$$

and

$$\sigma^2 = \text{Var}(X) = \frac{(\theta_2 - \theta_1)^2}{12}$$

- For the ND basketball team, what will be the expected value of points scored?

PDF and CDF

- Is the Uniform a proper PDF?
- What if we want to calculate $P(P \geq 75)$?
  - We could use integrals.
  - Or

$$F(c) = \frac{c - \theta_1}{\theta_2 - \theta_1}$$

- What is $P(P > 90)$?

Example: Nonameexample

The random variable X is known to be uniformly distributed between 10 and 20.

1. Show the graph of the PDF.
2. Compute $P(X < 15)$.
3. Compute $P(12 \leq X \leq 18)$.
4. Compute $\mathbb{E}(X)$ and $\text{Var}(X)$.

From Binomial To Normal

In the Limit...

- Poisson in Excel.
- Binomial in Excel.

Consider the purchase decision of the next three customers who enter the Martin Clothing Store. On the basis of past experience, the store manager estimates the probability that any one customer will make a purchase is 0.30.

1. What is the probability that two of the next three customers will make a purchase?

The Normal Distribution

- Motivation.

A r.v. X is said to have a *normal probability distribution* iff, for $\sigma > 0$ and $-\infty < \mu < \infty$, the density function of X is

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx \quad -\infty < x < \infty$$

If X is a normally distributed random variable with parameters μ and σ, then

$$\mathbb{E}(X) = \mu$$

and

$$\mathrm{Var}(X) = \sigma^2$$

- Importance of those *parameters*.
- What if we want to calculate the probability of an interval?
  - Bad news if you want a closed form solution.
  - Good news for you.

Some Properties

- Symmetry.
- Mode.
- What is the median?
- Skewness.
- Almost all area is between $(-3\sigma, 3\sigma)$.

The Standard Normal Distribution

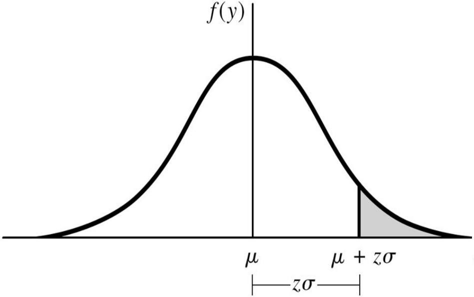
- Transforming the normal distribution. Remember the z-score?
- Reading probabilities in a table.

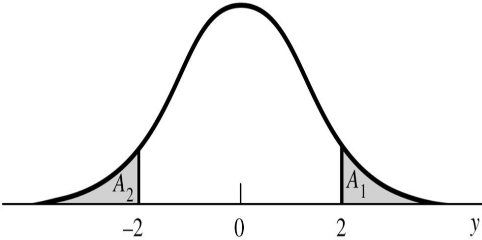
Let Z denote a normal random variable with mean 0 and standard deviation 1.

1. Find $P(Z > 2)$.
2. Find $P(-2 \leq Z \leq 2)$.
3. Find $P(0 \leq Z \leq 1.73)$.

$P(Z > 2)$

$P(-2 \leq Z \leq 2)$





Some Notation

- Mathematically:

The PDF for a Standard Normal Distribution is

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} dz \equiv \phi(z) \quad -\infty < z < \infty$$

- What is $\phi(z)$?
  - $\Phi(z)$ is the standard normal evaluated at z.

$$\Phi(a) = P(Z \leq a) = \int_{-\infty}^{a} \phi(z) dz$$

Three Types of Calculations

1. $P(Z \leq a) = \Phi(a)$
2. $P(Z > a) = 1 - \Phi(a)$
3. $P(a \leq Z \leq b) = \Phi(b) - \Phi(a)$

Less or Equal Than I: $P(Z \leq -0.98)$

0.09	0.08	0.07	0.06	0.05	0.04	0.03	0.02	0.01	0	Z
0.0559	0.0571	0.0582	0.0594	0.0606	0.0618	0.0630	0.0643	0.0655	0.0668	-1.50
0.0681	0.0694	0.0708	0.0721	0.0735	0.0749	0.0764	0.0778	0.0793	0.0808	-1.40
0.0823	0.0838	0.0853	0.0869	0.0885	0.0901	0.0918	0.0934	0.0951	0.0968	-1.30
0.0985	0.1003	0.1020	0.1038	0.1056	0.1075	0.1093	0.1112	0.1131	0.1151	-1.20
0.1170	0.1190	0.1210	0.1230	0.1251	0.1271	0.1292	0.1314	0.1335	0.1357	-1.10
0.1379	0.1401	0.1423	0.1446	0.1469	0.1492	0.1515	0.1539	0.1562	0.1587	-1.00
0.1611	0.1635	0.1660	0.1685	0.1711	0.1736	0.1762	0.1788	0.1814	0.1841	-0.90
0.1867	0.1894	0.1922	0.1949	0.1977	0.2005	0.2033	0.2061	0.2090	0.2119	-0.80
0.2148	0.2177	0.2206	0.2236	0.2266	0.2296	0.2327	0.2358	0.2389	0.2420	-0.70
0.2451	0.2483	0.2514	0.2546	0.2578	0.2611	0.2643	0.2676	0.2709	0.2743	-0.60
0.2776	0.2810	0.2843	0.2877	0.2912	0.2946	0.2981	0.3015	0.3050	0.3085	-0.50

Less or Equal Than II: $P(Z \leq 1.41)$

Z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.50	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.60	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.70	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.80	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.90	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.00	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.10	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.20	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.30	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.40	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.50	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441

Greater Than: $P(Z > 1.17)$

Z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.50	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.60	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.70	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.80	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.90	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.00	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.10	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.20	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.30	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.40	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.50	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441

Between: $P(0.1 \leq Z \leq 1.9)$

Z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
1.00	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.10	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.20	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.30	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.40	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.50	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.60	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.70	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.80	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.90	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.00	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817

Between: $P(0.1 \leq Z \leq 1.9)$

Z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.00	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.10	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.20	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.30	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.40	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.50	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.60	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.70	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.80	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.90	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.00	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621

Evaluating a Normal Distribution

- How to evaluate a normal distribution with different means and variances?
- Suppose $X \sim \mathcal{N}(\mu_x, \sigma_x^2)$.
  - What is $P(X \leq a)$?
- Going green: linear transformation of random variables.
  - Create a "new" random variable Z.
  - What are the mean and variance of Z? How is it distributed?

If X is normally distributed with mean μ_x and variance σ^2, then any linear combination of X is also normal.

- For instance if $Y = \beta_0 + \beta_1 X$.

Some Examples

If $X \sim \mathcal{N}(5, 4)$ what is $P(X < 7.5)$?

If $X \sim \mathcal{N}(12, 9)$ what is $P(X > 6)$?

If $X \sim \mathcal{N}(6, 5)$ what is $P(4 < X < 10)$?

Example: Mickey Mouse Taking a Test

To ride the "Space Mountain" at Walt Disney, guest must be greater than 44 inches tall. Suppose that the heights of all visitors to Disney World are normally distributed with a mean of 53 inches and a standard deviation of 6.

1. What fraction of visitors to Disney World can ride Space Mountain?

Time to complete an exam is normally distributed with a mean of 110 and a standard deviation of 20.

1. What is the fraction of students who will not finish within 120 minutes?

A Little Twist

To graduate with honors, you must be on the top 5 percent (*Summa Cum Laude*), 15 percent (*Magna Cum Laude*), or 30 percent (*Cum Laude*). Suppose GPA is distributed normally with a mean of 2.6 and a standard deviation of 0.65.

1. What GPA will you need to graduate Cum Laude?

Grear Tires is considering a guarantee that will provide a discount on replacement tires if the original tires do not provide the guaranteed mileage. Assume that the mean tire mileage is $\mu = 36,500$ and the standard deviation is $\sigma = 5,000$.

1. What should the guarantee mileage be if the firms wants no more than 10% of the tires to be eligible for the discount guarantee?

Excel and Probability Table

- Poisson in Excel.
- Binomial in Excel.

Consider the purchase decision of the next three customers who enter the Martin Clothing Store. On the basis of past experience, the store manager estimates the probability that any one customer will make a purchase is 0.30.

1. What is the probability that two of the next three customers will make a purchase?

- Going to the limit.

Normal Approximation of Binomial Probabilities

- Why?
- The $np \geq 5$ and $n(1-p) \geq 5$ rules.
- How would our 'Normal' be distributed?

A particular company has a history of making errors in 10% of its invoices. A sample of 100 invoices has been taken.

1. What is the probability that 12 invoices contain errors?
2. What is the probability that 13 or fewer errors in the sample of 100 invoices?

- The **Continuity Correction Factor**.

Exponential Distribution

Exponential Probability Distribution

- Examples.

A r.v. X is said to have a *exponential distribution with parameter* β iff the density function of X is

$$f(x) = \begin{cases} \frac{1}{\beta} e^{-\frac{x}{\beta}} & \text{for } x \geq 0 \\ 0 & \text{elsewhere} \end{cases}$$

with $\beta > 0$, and β is the expected value of X.

- Is this a valid distribution?
- What is the CDF?

Example: Car I

The time to when a new car from a certain manufacture will fail is described by an exponential distribution with a mean of 6 years.

1. What is the probability that the car will die in just three years?
2. What is the median time to failure?

Example: Car II

Suppose a 'lemon' is a car that dies within 3 years. Five people on your dorm all purchase the same type of car. Assume $\beta = 6$.

1. What is the probability that all 5 cars will be lemons?

Exponential and Poisson

- Relationship between Exponential and Poisson.

If arrivals follow a Poisson distribution, the time between arrivals must follow an exponential distribution.

- Example:

The number of graduate students that arrive at Burger King during one hour is described by a Poisson probability distribution with a mean of 10 graduate students per hour.

1. How is the Poisson that describes this event?
2. How is the corresponding exponential distribution that describe the time between arrivals?