Probability Density Functions

A high level overview and numerical sampling strategies

Andrew Siegel

April 3, 2017

Outline

Outline

PDFs

Basics

- In an experiment, we can describe the unknown outcome with a random variable X. Though the particular outcome is unknown, the probability of any outcome can be described.
- For a random variable X, the *cumlative distribution function* (or CDF) is the probability that the value of X is less than some value x:

$$F(x) = P\{X < x\}$$

- The *probability mass function* $p(x_i)$ describes the probability for a countably-infinite number of discrete outcomes $x_1, x_2, \ldots$ such that

$$p(x_i) = P\{X = x_i\}$$

and has the property:

$$\sum_{i=1}^{n} p(x_i) = 1$$

PDFs

Basics

- For a continuously-valued random variable X, The probability density function $f(x)$ describes the probability for some range C of continuously-valued outcomes:

$$\int_C f(x)\,dx = P\{X \in C\}$$

- Note that for any single value X, the probability of that value is identically zero (why?). We can express the PDF for for an infinitesimally-small range ϵ as:

$$P\left\{a - \frac{\epsilon}{2} \leq X \leq a + \frac{\epsilon}{2}\right\} = \int_{a-\frac{\epsilon}{2}}^{a+\frac{\epsilon}{2}} f(x)\,dx \approx \epsilon f(a)$$

Outline

PDFs

Joint PDFs

- For two discrete random variables X and Y, the *joint probability mass function* $p(x, y)$ is described as:

$$p(x, y) = P\{X = x, Y = y\}$$

- Likewise, for two continous random variables X and Y, the *joint probability density function* is described as:

$$\int_D \int_C f(x, y)\, dx\, dy = P\{X \in C, Y \in D\}$$

- The variables X and Y are *independent* if:

$$f(x, y) = f_X(x) f_Y(y)$$

where

$$f_X(x) = \int_{-\infty}^{\infty} f(x, y) dy; f_Y(y) = \int_{-\infty}^{\infty} f(x, y) dx$$

$f_X(x)$ and $f_Y(y)$ are known as the *marginal PDFs*.

PDFs

Marginal Distribution Example

Y \ X	x_1	x_2	x_3	x_4	$p_y(Y)\downarrow$
y_1	$4/32$	$2/32$	$1/32$	$1/32$	$8/32$
y_2	$2/32$	$4/32$	$1/32$	$1/32$	$8/32$
y_3	$2/32$	$2/32$	$2/32$	$2/32$	$8/32$
y_4	$8/32$	0	0	0	$8/32$
$p_x(X) \rightarrow$	$16/32$	$8/32$	$4/32$	$4/32$	$32/32$

Outline

PDFs

Uniform Distribtion

- A random variable X is said to be uniformly distributed over the interval (a, b) if it has a pdf given by:

$$f(x) = \begin{cases} \frac{1}{b-a} \text{ if } a < x < b \\ 0 \text{ otherwise} \end{cases}$$

- The mean is thus:

$$E[X] = \frac{1}{b-a} \int_a^b x dx = \frac{b^2 - a^2}{2(b-a)} = \frac{b+a}{2}$$

- and the second moment is:

$$E[X^2] = \frac{1}{b-a} \int_a^b x^2 dx = \frac{b^3 - a^3}{3(b-a)} = \frac{a^2 + b^2 + ab}{3}$$

PDFs

Uniform Distribtion, cont.

- and the variance is thus

$$Var[X] = \frac{1}{12}(b-a)^2$$

- and the CDF is

$$F(x) = P\{X \leq x\} = \int_a^x \frac{1}{b-a}dx = \frac{x-a}{b-a}$$

PDFs

Gaussian Distribution

- A random variable is said to be normally distributed with mean μ and variance σ^2 if its pdf is given by:

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

- The expected value is then

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

- and the variance is

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

Outline

- Markov Inequality

$$P\{X \geq a\} \leq \frac{E[X]}{a}$$

- Proof

$$\begin{aligned}
E[X] &= \int_0^{\infty} xf(x)dx \\
&= \int_0^{a} xf(x)dx + \int_a^{\infty} xf(x)dx \\
&\geq \int_a^{\infty} xf(x)dx \\
&\geq \int_a^{\infty} af(x)dx \\
&= a\int_a^{\infty} f(x)dx \\
&= aP\{X \geq a\}
\end{aligned}$$

Markov Inequality

Intuition

- Imagine a 6-sided fair die with $E[X] = 3.5$
- Let's use Markov to bound e.g. $P(x \geq 6)$.
- Markov: $P\{X \geq a\} \leq \frac{E[X]}{a}$
- In this case $P(x \geq 6) \leq \frac{3.5}{6} = \frac{7}{12}$
- Imagine this is not the case, and that $P(x \geq 6) > \frac{7}{12}$

$$E[X] = 1P(X = 1) + 2P(X = 2) + \cdots + 6P(X = 6) \geq 6P(X = 6)$$

- But if $P(X = 6) > \frac{7}{12}$ then $E[X] > 3.5$.
- Contradiction!

Weak Law

Chebyshev Inequality

- Chebyshev Inequality says that for any value $k > 0$

$$P\{|X-\mu| \geq k\} \leq \frac{\sigma^2}{k^2}$$

That is, most points are close to the mean.

- Proof. Since $(X-\mu)^2$ is a nonnegative random variable we can apply Markov's inequality with $a = k^2$:

$$P\{(X-\mu)^2 \geq k^2\} \leq \frac{E[(X-\mu)^2]}{k^2}$$

- But since $|X-\mu|^2 \geq k^2$ if and only if $|X-\mu| \geq k$, then the distributions $P\{|X-\mu| \geq k\}$ and $P\{(X-\mu)^2 \geq k^2\}$ are **identical and**

$$P\{|X-\mu| \geq k\} \leq \frac{E[(X-\mu)^2]}{k^2} = \frac{\sigma^2}{k^2}$$

Weak Law of Large Numbers

Weak Law Proof

- Weak Law of Large Numbers: Let $X_1, X_2, \cdots$, be a sequence of i.i.d. random variables each with mean μ. Then for any $\epsilon > 0$

$$P\left\{\left|\frac{X_1 + \cdots + X_n}{n} - \mu\right| > \epsilon\right\} \to 0 \text{ as n} \to \infty$$

- Proof: Since

$$E\left[\frac{X_1 + \cdots + X_n}{n}\right] = \mu \text{ and } Var\left(\frac{X_1 + \cdots + X_n}{n}\right) = \frac{\sigma^2}{n}$$

then since $Var(\frac{X_1+\cdots+X_n}{n}) = \frac{\sigma^2}{n}$ we have

$$P\left\{\left|\frac{X_1 + \cdots + X_n}{n} - \mu\right| > \epsilon\right\} \leq \frac{\sigma^2}{n\epsilon^2}$$

Outline

PDFs

Central Limit Theorem

- One of the most important theorems in all of statistics.
- Has many different forms, will look at others later
- Start with a random sample of size n $X_1, X_2, \cdots, X_n$ where the X_i are iid with mean μ and variance σ^2.
- Then, consider the sample mean of the X_i

$$S_n = \frac{X_1 + X_2 + \cdots + X_n}{n}$$

- The *Law of Large Numbers* says that the sequence S_n *converges in probability* to the true mean μ as $n \to \infty$
- *The Central Limit Theorem* says, roughly, that the sample means are normally distributed as $n \to \infty$ **independent of the distribution of the individual X_i.**

PDFs

Central Limit Theorem, cont.

- A more precise statement is the *Lindeberg–Levy CLT*.
- We ask at what rate does $\overline{x}_n \to \mu$
- It can be easily shown that $|\overline{x}_n - \mu|$ tends to zero in distribution at the rate $\frac{1}{\sqrt{n}}$.
- This leads to Lindeberg-Levy CLT, which gives a non-degenerate form of the limiting distribution. It is one form of the Central Limit Theorem:

$$\sqrt{n}\left(\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) - \mu\right) \to N(0, \sigma^2)$$

- The difference between the sample mean and true mean when multiplied by the square root of the number of samples tends to a normal distribution

Outline

Sampling from an Arbitrary PDF

Inverse Transform Method

- How do you sample a random value from an arbitrary PDF or probability mass function?
- The *inverse transform method* is a simple technique appropriate for both discrete or continuous random variables.
- First consider the continuous case. Suppose you want to generate a random variable X with an arbitrary CDF, $F(x)$.
- Assuming you have a way to generate $U(0,1)$, a uniform random number between 0 and 1, then the following algorithm applies:

Sampling from an Arbitrary PDF

Inverse Transform Method

- Algorithm
  1. Generate U from $U(0,1)$
  2. Find a value of X such that $F(X) = U$
- That is, we want to solve $X = F^{-1}(U)$.
- Note that there are some limitations: first, if the CDF is given as a continuous function it must be invertible, which is not always the case
- If the CFD is discrete we interpret step (2) to mean the first index of cdf such that $U \geq cdf$.
- Also note that non-ivertible functions can be discretized even though this process is often expensive. Other approaches may be superior (see Rejection sampling).

Sampling from an Arbitrary PDF

Inverse Transform Method

- Proof

$$
\begin{aligned}
P\{X \leq x_0\} &= P\{F^{-1}(U) \leq x_0\} \\
&= P\{F(F^{-1}(U)) \leq F(x_0)\} && \text{since } F(x) \text{ is increasing} \\
&= P\{U \leq F(x_0)\} \\
&= F(x_0) && \text{since } U \text{ is uniform}
\end{aligned}
$$

Sampling from an Arbitrary PDF

Inverse Transform Method

- Example: Consider the PDF

$$p(x) = e^{-x} \quad 0 < x < \infty$$

- The CDF is

$$F(x) = 1 - e^{-x}$$

.

- The inverse, which yields a random variable X, is

$$F^{-1}(U) = -\log(1 - U) = X$$

.

- A discrete form of the algorithm is extremely useful for tabulated functions or those that cannot be inverted. See Ross and course demos.

Outline

Sampling from an Arbitrary distribution

Rejection Sampling

- Goal is still to sample a random variable X from an arbitrary PDF $f(x)$.
- uppose you have a method of sampling from some other PDF, $g(x)$ such that, for some constant M:

$$f(x) \leq Mg(x)$$

- Then the following algorithm applies:
  1. Sample X from $g(x)$
  2. Calculate α, the probability of accepting X:

$$\alpha = \frac{f(x)}{Mg(x)}$$

  3. Sample U from $U(0,1)$
  4. If $\alpha \geq U$, then accept the value of X. If not, reject X and repeat.

Sampling from an Arbitrary distribution

Rejection Sampling

- Rejection Sampling is inefficient if $f(x)$ and $g(x)$ are not sufficiently similar, since the chances of accepting X are low.
- If $f(x)$ and $g(x)$ are "similar", then α is closer to 1, and the method is likely to be efficient
- In practice rejection sampling is very useful if you aren't overly concerned with performance since it is completely general and very simple to code.
- See course examples