Expectations

Probability calculus / Adv Stat I

Prof. Dr. Matei Demetrescu

The cdf is very informative

The cdf (or pdf) describes *an entire* distribution.

This may be overkill at times,

and we may want to focus on specific characteristics.

We often resort to average characteristics of possible outcomes,

... so today we formalize this notion (and also give it a new name).

Today's outline

Expectations

1. Expectation of a random variable
2. Properties of the expectation operator
3. Representing discrete pdfs via expectations
4. Up next

Outline

1. Expectation of a random variable
2. Properties of the expectation operator
3. Representing discrete pdfs via expectations
4. Up next

The nexus

The expected value, or expectation, of a random variable represents its probability-weighted average value;

It gives a measure of the location of X (the center of gravity of its pdf).

Definition (Expectation; discrete case)

The expected value of a discrete random variable exists, and is defined by

$$\mathrm{E}(X) = \sum_{x \in \mathrm{R}(X)} x \cdot f(x), \quad \text{iff} \quad \sum_{x \in \mathrm{R}(X)} |x \cdot f(x)| = \sum_{x \in \mathrm{R}(X)} |x| \cdot f(x) < \infty.$$

- The existence condition ensures that the sum $\sum_{x \in \mathrm{R}(X)} x f(x)$ defining the expectation is absolutely convergent.
- The condition is sometimes called integrability of X.

Worth thinking about

- Thanks to the triangle inequality, absolute convergence implies standard convergence:

$$\sum_{x\in \mathrm{R}(X)} |x|\cdot f\left(x\right)<\infty \quad \Rightarrow \quad \Big|\sum_{x\in \mathrm{R}(X)} x\cdot f\left(x\right)\Big|<\infty.$$

  such that the sum defining the expectation is finite and exists.
- Also, if $\mathrm{R}(X)$ is finite and $|x|<\infty\ \forall\ x\in \mathrm{R}(X)$, then $\sum_{x\in \mathrm{R}(X)} |x|\cdot f\left(x\right)<\infty$ automatically.
- But if $\mathrm{R}(X)$ is countably infinite there is no guarantee that $\sum_{x\in \mathrm{R}(X)} |x|\cdot f\left(x\right)<\infty$.

Example

Consider a discrete random variable with pdf

$$f(x_k) = \frac{1}{2^k} \quad \text{with} \quad \mathrm{R}(X) = \left\{ x_k = (-1)^k \frac{2^k}{k}, k = 1, 2, \dots \right\}.$$

The sum defining the expectation is

$$\begin{aligned} \sum_{k=1}^{\infty} x_k f(x_k) &= \sum_{k=1}^{\infty} \frac{(-1)^k}{k} = -\sum_{k=1}^{\infty} \frac{(-1)^{k-1}}{k} 1^k \\ &= -\ln(1+1). \qquad \left[\sum_{k=1}^{\infty} \frac{(-1)^{k-1}}{k} x^k \Big|_{x \in (-1,1]} = \ln(1+x) \right] \end{aligned}$$

Thus, the sum is convergent, but not absolutely convergent since

$$\sum_{k=1}^{\infty} |x_k| f(x_k) = \sum_{k=1}^{\infty} \frac{1}{k} = 1 + \frac{1}{2} + \underbrace{\frac{1}{3} + \frac{1}{4}}_{>1/2} + \underbrace{\frac{1}{5} + \cdots + \frac{1}{8}}_{>1/2} + \cdots = \infty.$$

Moving on to integrals

Definition (Expectation; continuous case)

The expected value of a continuous random variable exists, and is defined by

$$\mathrm{E}(X) = \int_{-\infty}^{\infty} x \cdot f(x) \, \mathrm{d}x, \quad \text{iff} \quad \int_{-\infty}^{\infty} |x| \cdot f(x) \, \mathrm{d}x < \infty.$$

The existence condition is necessary to ensure that the improper Riemann integral $\int_{-\infty}^{\infty} x \cdot f(x) \, \mathrm{d}x$ (and hence the expectation) exists.

Theorem (3.1)

If $|x| < c \; \forall \; x \in \mathrm{R}(X)$, *for some choice of* $c \in (0, \infty)$. *Then* $\mathrm{E}(X)$ *exists.*

The Cauchy (counter) example

Consider a random variable with pdf

$$f(x) = \frac{1}{\pi}\frac{1}{1+x^2}, \qquad -\infty < x < \infty \qquad \text{(Cauchy distribution)}.$$

Write now

$$\int_{-\infty}^{\infty} |x| f(x) \mathrm{d}x = \int_{-\infty}^{\infty} \frac{|x|}{\pi}\frac{1}{1+x^2}\mathrm{d}x = \frac{2}{\pi}\int_0^{\infty}\frac{x}{1+x^2}\mathrm{d}x,$$

For any positive number a we obtain

$$\int_0^a \frac{x}{1+x^2}\mathrm{d}x = \left[\frac{\ln(1+x^2)}{2}\right]_{x=0}^{x=a} = \frac{\ln(1+a^2)}{2}.$$

Thus,

$$\int_{-\infty}^{\infty} |x| f(x) \mathrm{d}x = \lim_{a\to\infty}\frac{2}{\pi}\int_0^a \frac{x}{1+x^2}\mathrm{d}x = \frac{1}{\pi}\lim_{a\to\infty}\ln(1+a^2) = \infty.$$

Outline

1. Expectation of a random variable
2. Properties of the expectation operator
3. Representing discrete pdfs via expectations
4. Up next

Expectation of a function of random variables

We sometimes need to work with transformations of RVs, $Y = g(X)$.

If we only need $\mathrm{E}(Y)$, we don't have to derive the pdf of Y.[1]

Theorem (3.2)

Let X be a random variable with pdf $f(x)$. Then the expectation of random variable $Y = g(X)$ is given by

$$\mathrm{E}(g(X)) = \begin{cases} \displaystyle\sum_{x \in \mathrm{R}(X)} g(x)\, f(x) & \textit{(discrete)} \\ \displaystyle\int_{-\infty}^{\infty} g(x)\, f(x)\, \mathrm{d}x & \textit{(continuous).} \end{cases}$$

[1] Fortunately.

Probability and expectation

An application of the theorem is that the expectation of an indicator function equals the probability of the set being indicated:

Example

Let X be a variable with pdf f and recall that $\mathbb{I}_A(x) = \begin{cases} 1 & x \in A \\ 0 & \text{else} \end{cases}$.

Then,

$$\mathrm{E}\left(\mathbb{I}_A(X)\right) = \begin{cases} \displaystyle\sum_{x \in \mathrm{R}(X)} \mathbb{I}_A(x) \cdot f(x) = \sum_{x \in A} f(x) = \mathrm{P}(X \in A) & \text{(discrete)} \\ \displaystyle\int_{x \in \mathrm{R}(X)} \mathbb{I}_A(x) \cdot f(x) \mathrm{d}x = \int_{x \in A} f(x) \mathrm{d}x = \mathrm{P}(X \in A) & \text{(cont).} \end{cases}$$

Markov's inequality

Theorem (3.11 (Markov's inequality))

Let X be a random variable with pdf f, and let g be a nonnegative function of X. Then

$$\mathrm{P}(g(X) \geq a) \leq \frac{\mathrm{E}(g(X))}{a} \qquad \text{for any } a > 0.$$

Note that a should be large to have nontrivial bounds, though.

Can we say anything about $\mathrm{E}(g(X))$ in relation to $\mathrm{E}(X)$?

Theorem (3.3 (Jensen's Inequality))

Let X be a non-degenerate random variable with expectation $\mathrm{E}(X)$, and let g be a continuous function on an open interval I containing $\mathrm{R}(X)$ (that is $\mathrm{R}(X) \subseteq I$).

$$\begin{aligned} &\text{If } g \text{ is convex on } I, && \text{then } \mathrm{E}(g(X)) \geq g(\mathrm{E}(X)); \\ &\text{if } g \text{ is strictly convex on } I, && \text{then } \mathrm{E}(g(X)) > g(\mathrm{E}(X)). \end{aligned}$$

Jensen's Inequality also applies to concave functions...

One immediate application of Jensen's Inequality shows that

$$\mathrm{E}(X^2) \geq (\mathrm{E}(X))^2, \qquad \text{since } g(x) = x^2 \text{ is convex.}$$

Note that this implies that $\mathrm{Var}(X) = \mathrm{E}(X^2) - (\mathrm{E}(X))^2 \geq 0.$[2]

[2] In case you have not met the variance before: we introduce it next week formally.

Some more properties of the expectation

Theorem (3.4)

If c is a constant, then $\mathrm{E}(c) = c$.

Theorem (3.5)

If c is a constant, then $\mathrm{E}(cX) = c\,\mathrm{E}(X)$.

Theorem (3.6)

$\mathrm{E}\left(\sum_{i=1}^{k} g_i(X)\right) = \sum_{i=1}^{k} \mathrm{E}(g_i(X))$.

Corollary (3.1)

$\mathrm{E}(a + bX) = a + b\,\mathrm{E}(X)$.

Outline

1. Expectation of a random variable
2. Properties of the expectation operator
3. Representing discrete pdfs via expectations
4. Up next

Recall: the Riemann integral

Definition

A function g is said to be Riemann-integrable on an interval $[a, b]$ if, for any partition $a = x_0, \ldots, x_n = b$ of the interval $[a, b]$, the limit

$$\lim_{\max|x_i - x_{i-1}| \to 0} \sum_{i=1}^{n} g(\xi_i)\,(x_i - x_{i-1})$$

exists and is finite for any $\xi_i \in [x_i, x_{i-1}]$.

- The limit is then denoted by $\int_a^b g(x)\mathrm{d}x$
- Piecewise continuity is sufficient for (Riemann) integrability
- The (Riemann) integral is "the area under the curve"

A generalization of Riemann integration

Definition

A function g is said to be Stieltjes-integrable on an interval $[a, b]$ with integrator F if, for any partition $a = x_0, \ldots, x_n = b$ of the interval $[a, b]$, the limit

$$\lim_{\max|x_i - x_{i-1}| \to 0} \sum_{i=1}^{n} g(\xi_i) \left(F(x_i) - F(x_{i-1}) \right)$$

exists and is finite for any $\xi_i \in [x_i, x_{i-1}]$.

- The limit is then denoted by $\int_a^b g(x) \mathrm{d}F(x)$
- Piecewise continuity of g and monotonicity of F are sufficient for Stieltjes integrability, provided that discontinuities of g and F are not common.
- Improper integrals and integrals over unions of intervals are treated the usual (Riemann) way

Equivalence

The Stieltjes and Riemann integrals are closely related

- In fact, if $F(x)$ is linear, they are (more or less) the same.
- Moreover, if F is smooth, $\int_A g(x)\mathrm{d}F(x) = \int_A g(x)F'(x)\mathrm{d}x$
- This is relevant for distributions:

Example

Let f be the pdf of a continuous random variable X and F (f) the associated cdf (pdf). Then,

$$\mathrm{E}\left(g(X)\right) = \int_{\mathrm{R}(X)} g(x)f(x)\mathrm{d}x = \int_{\mathrm{R}(X)} g(x)\mathrm{d}F(x).$$

Properties at a glance

Theorem

Let $g : [a, b] \to \mathbb{R}$ be Stietltjes integrable w.r.t. right-continuous F. Then,

1. *Linearity: for $A, B \in \mathbb{R}$,*

$$\int_a^b \left(Ag_1(x) + Bg_2(x)\right) \mathrm{d}F(x) = A \int_a^b g_1(x)\mathrm{d}F(x) + B \int_a^b g_2(x)\mathrm{d}F(x)$$

$$\int_a^b g(x)\mathrm{d}\left(AF_1(x) + BF_2(x)\right) = A \int_a^b g(x)\mathrm{d}F_1(x) + B \int_a^b g(x)\mathrm{d}F_2(x)$$

$$\int_a^b g(x)\mathrm{d}F(x) = \int_a^c g(x)\mathrm{d}F(x) + \int_c^b g(x)\mathrm{d}F(x) \quad \text{where } c \in (a, b)\,.$$

2. *Integration by parts: $\int_a^b g(x)\mathrm{d}F(x) = g(x)F(x)|_a^b - \int_a^b F(x)\mathrm{d}g(x)$.*
3. *Equivalence with Riemann integral when F is smooth.*
4. *Change of variables: $\int_c^d g(h(y))\mathrm{d}F(h(y)) = \int_{h(c)}^{h(d)} g(x)\mathrm{d}F(x)$.*

The Stieltjes integral is more flexible

The integrator F does not have to be continuous!

Lemma

Let F be piecewise smooth, right-continuous with a jump discontinuity at $x = x_0 \in [a, b]$, and g piecewise continuous, continuous at x_0. Then,

$$\begin{aligned}\int_a^b g(x)\mathrm{d}F(x) &= \lim_{c \nearrow x_0} \int_a^c g(x)\mathrm{d}F(x) + \int_{x_0}^b g(x)\mathrm{d}F(x) \\ &\quad + g(x_0)\left(F(x_0+) - F(x_0-)\right)\end{aligned}$$

where $F(x_0+)$ ($F(x_0-)$) stands for the limit of F at x_0 to the right (to the left).

Unifying discrete and continuous distributions

Example

Take the two-point distribution given by

$$\mathrm{P}(X=0)=1-p \quad \text{and} \quad \mathrm{P}(X=1)=p$$

(the Bernoulli distribution with success probability p), with expectation

$$\mathrm{E}(X)=0\cdot(1-p)+1\cdot p=p.$$

Its cdf is

$$F(x)=(1-p)\mathbb{I}(x\geq 0)+p\mathbb{I}(x\geq 1),$$

and the above lemma indicates that

$$\int_{-\infty}^{\infty} x\mathrm{d}F(x)=p=\mathrm{E}(X).$$

A unified notation

Recall: if F is smooth,

$$\int_a^b g(x)\mathrm{d}F(x) = \int_a^b g(x)F'(x)\mathrm{d}x.$$

Then...

- The Stieltjes integral on the l.h.s. exists for discontinuous F as well.
- May exploit the equality to define a "derivative" of F at its jumps.
- Focus to this end on piecewise smooth F with one finite jump at $x_0 \in (a, b)$.

Jumps I

Split F in smooth and nonsmooth components,

$$F(x) = \tilde{F}(x) + (F(x_0+) - F(x_0-))\, H(x - x_0)$$

where $\tilde{F}$ is smooth and $H(x)$ is a jump function at 0,

$$H(x) = \begin{cases} 0 & x < 0 \\ 1 & x \geq 0 \end{cases}.$$

With $C = F(x_0+) - F(x_0-)$, the previous Lemma delivers

$$\int_a^b g(x)\mathrm{d}F(x) = \int_a^b g(x)\mathrm{d}\tilde{F}(x) + C \int_a^b g(x)\mathrm{d}H(x - x_0).$$

Jumps II

Would we be able to differentiate H,

$$\int_a^b g(x)\mathrm{d}F(x) = \int_a^b g(x)\tilde{F}'(x)\mathrm{d}x + C\int_a^b g(x)H'(x-x_0)\mathrm{d}x.$$

At the same time (see previous Lemma),

$$\begin{aligned}\int_a^b g(x)\mathrm{d}F(x) &= \int_a^{x_0} g(x)\mathrm{d}F(x) + \int_{x_0}^b g(x)\mathrm{d}F(x) + Cg(x_0) \\ &= \int_a^b g(x)\tilde{F}'(x)\mathrm{d}x + Cg(x_0)\end{aligned}$$

since $\mathrm{d}\tilde{F} = \mathrm{d}F$ for $x < x_0$ and $x > x_0$. Therefore, we should have

$$\int_a^b g(x)H'(x-x_0)\mathrm{d}x = g\left(x_0\right).$$

At the same time, $H' = 0$ for all $x \neq 0$, but H' is undefined at $x = 0$.

The δ function

We now have all the ingredients we need:

Definition (Dirac's δ)

The generalized function $\delta(x)$ satisfying

1. $\delta(x) = 0$ for all $x \neq 0$ and
2. $\int_{-\infty}^{\infty} g(x)\delta(x)\mathrm{d}x = g\,(0)$

is called the Dirac's δ, and we write $\delta(x) = H'(x)$.

Dirac's δ can also be seen as the pdf of the limit of convergence in probability to a constant.

Discrete pdfs

We may now write for any discrete pdf

$$f(x) = f(x)\mathbb{I}_{R(X)}(x)$$

with the interpretation that $f(x) = \mathrm{P}(X = x)$[3] in a more intuitive way,

$$f(x) = \sum_{x_0 \in \mathrm{R}(X)} \mathrm{P}(X = x_0)\, \delta(x - x_0),$$

one that also allows for nice integration.

The (generalized) pdf of a point mass distribution (i.e. cdf $H(x - x_0)$) is thus $\delta(x - x_0)$.

[3] But not with the properties of a "proper" density function, since, in the Riemann world, $\int_{\mathrm{R}(X)} f(x)\mathbb{I}_A(x)\mathrm{d}x = 0$ for discrete sets A.

Outline

1. Expectation of a random variable
2. Properties of the expectation operator
3. Representing discrete pdfs via expectations
4. Up next

Coming up

Moments and other functionals