Moments of Random Variables

Probability calculus / Adv Stat I

Prof. Dr. Matei Demetrescu

Summarizing a distribution?

Expectations allow one to discuss "average", or "typical", outcomes in a rigorous manner.

We motivated this by the need to analyze distributions using selected average characteristics rather than the pdf or cdf.

- We note that such averages do not replace the cdf/pdf,
- ... they rather help us grasp various aspects of cdfs/pdfs more intuitively.

Today, we'll get more specific and target particular aspects of probability distributions by focusing on specific expectations.

Moments of Random Variables

Outline

Special expectations

Moments are expectations of power functions of random variables. They can be used to measure certain characteristics of given distributions.

We distinguish between non-central and central moments.

Definition (rth non-central moment)

Let X be a random variable with pdf $f(x)$. Then the rth non-central moment of X, denoted by μ_r', is defined as

$$\mu_r' = \mathrm{E}\left(X^r\right) = \begin{cases} \displaystyle\sum_{x \in \mathrm{R}(X)} x^r f\left(x\right) & \text{(discrete)} \\ \displaystyle\int_{-\infty}^{\infty} x^r f\left(x\right) \mathrm{d}x & \text{(continuous).} \end{cases}$$

Remarks

- Note that $\mu_0' = \mathrm{E}\left(X^0\right) = 1$.
- The first non-central moment is simply the expectation (also called the mean) of the random variable, that is $\mu_1' = \mathrm{E}\left(X\right)$; it is often denoted by μ.
- The mean is informative about the location of a distribution.
- Central moments aim at a location-free characterization of a distribution.

... and even more special ones

Definition (rth central moment)

Let X be a random variable with pdf $f(x)$. Then the rth central moment of X, denoted by μ_r, is defined as

$$\mu_r = \mathrm{E}\left((X-\mu)^r\right) = \begin{cases} \displaystyle\sum_{x \in \mathrm{R}(X)} (x-\mu)^r f(x) & \text{(discrete)} \\ \displaystyle\int_{-\infty}^{\infty} (x-\mu)^r f(x)\,\mathrm{d}x & \text{(continuous).} \end{cases}$$

- Note that $\mu_0 = \mathrm{E}((X-\mu)^0) = 1$, and $\mu_1 = \mathrm{E}(X-\mu) = 0$.
- The second central moment is commonly known as the variance.

Existence of Moments

Theorem (3.12)

If $\mathrm{E}(|X|^r)$ *exists for an* $r > 0$*, then* $\mathrm{E}(|X|^s)$ *exists* $\forall s \in [0, r]$.

The theorem implies, that if $\mathrm{E}(|X|^r)$ does not exist, then necessarily $\mathrm{E}(|X|^s)$ cannot exist for $s > r$.

Also, it noncentral moments of order $r \geq 1$ exist, so do noncentral ones.

Also,

Theorem (3.13)

If $\mathrm{E}(|Y - \mu|^r)$ *exists for an* $r > 0$*, then* $\mathrm{E}(|Y - \mu|^s)$ *exists* $\forall s \in [0, r]$.

Example

Consider the pdf

$$f(x) = \frac{2}{(x+1)^3}\mathbb{I}_{[0,\infty)}(x).$$

Examine $\mathrm{E}(X^\alpha)$, i.e.

$$\mathrm{E}\,X^\alpha = \int_0^\infty \frac{x^\alpha 2}{(x+1)^3}\mathrm{d}x = 2\int_1^\infty (y-1)^\alpha y^{-3}\mathrm{d}y,$$

(obtained by substituting $y = x + 1$, so that $y - 1 = x$ and $\mathrm{d}y = \mathrm{d}x$). If $\alpha = 2$, we get

$$\mathrm{E}(X^2) = 2\lim_{b\to\infty}\left[\ln(y) + 2y^{-1} - \frac{1}{2}y^{-2}\right]_{y=1}^{y=b} = \infty.$$

Thus, $\mathrm{E}(X^2)$ does not exist. By the theorem, moments of order larger than 2 also do not exist.

Outline

Variance and standard deviation

Definition

The **variance** of a random variable X is the 2nd central moment, $\mathrm{Var}(X) = \mathrm{E}\left((X-\mu)^2\right)$, and will be denoted by the symbol σ^2.
The non-negative square root of $\mathrm{Var}(X)$ is the **standard deviation** of X and will be denoted by the symbol σ.

- The variance and standard deviation are measures of the dispersion of a distribution around the mean.
- We have a simple connection to noncentral moments, $\mathrm{E}\left((X-\mu)^2\right) = \mathrm{E}(X^2) - \mu^2$.

A special case

Corollary (3.3 (Chebyshev's inequality))

$$\mathrm{P}(|X-\mu| \geq k\sigma) \leq \frac{1}{k^2} \quad \textit{for} \quad k > 0.$$

Equivalently, $\mathrm{P}(|X-\mu| \geq c) \leq \frac{\sigma^2}{c^2}$.

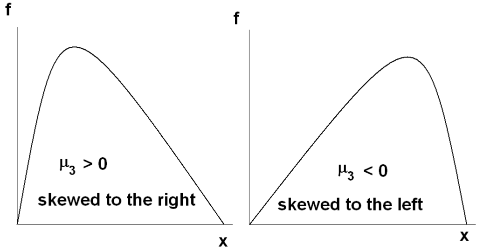
If $\sigma^2 \to 0$, then the distribution of X "collapses" to that of μ.

Further moments

Definition (Symmetry of a pdf)

The pdf f is said to be symmetric around μ iff
$f(\mu+\delta) = f(\mu-\delta)$ for any $\delta > 0$. Otherwise f is said to be skewed.

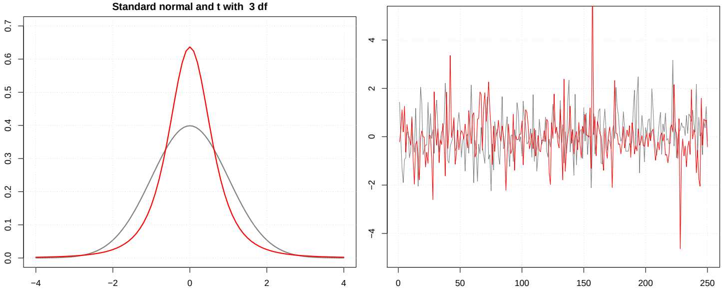
- A symmetric pdf has necessarily $\mu_3 = \mathrm{E}((X-\mu)^3) = 0$.
  (In fact, $\mu_{2K+1} = 0$ under symmetry.)
- For $\mu_3 > 0$ $(\mu_3 < 0)$ the pdf said to be skewed to the right (left).



The kurtosis

A distribution is said to have fat tails if it tends to generate too many large outcomes away from the bulk of the distribution.

To put things in perspective, recall the (standardized) t distribution, compared to the standard normal.



Relation between central and noncentral moments

Given noncentral moments we may derive the central ones, since

$$\mu_r = \mathrm{E}\left((X-\mu)^r\right) = \sum_{i=0}^{r} \binom{r}{i} (-1)^i \, \mathrm{E}(X^{r-i})\mu^i.$$

E.g. for the central 4th order moment we have

$$\mu_4 = \mathrm{E}(X^4) - 4\mu \, \mathrm{E}(X^3) + 6\mu^2 \, \mathrm{E}(X^2) - 3\mu^4.$$

(with μ the somewhat misleading but very common notation for $\mu_1' = \mathrm{E}(X)$.)

Mnemotechnics

The Moment-Generating Function (MGF) can be used to determine moments of a random variable.

Definition (Moment-Generating Function)

The MGF of a random variable X, denoted by $M_X(t)$, is

$$M_X(t) = \mathrm{E}\left(e^{tX}\right) = \begin{cases} \sum\limits_{x \in \mathrm{R}(X)} e^{tx} f(x) & \text{(discrete)} \\ \int_{-\infty}^{\infty} e^{tx} f(x) \, \mathrm{d}x & \text{(continuous),} \end{cases}$$

provided that the expectation exists for t in some neighborhood of 0. That is, there exists an $h > 0$ such that $\mathrm{E}\left(e^{tX}\right)$ exists $\forall\ t \in (-h, h)$.

The main practical use of the MGF is not to generate moments, but to help in characterizing a distribution.

The first use

The condition that $M_X(t)$ be defined $\forall\ t \in (-h, h)$ is a technical condition ensuring that $M_X(t)$ is differentiable at the point $t = 0$.

Differentiability is useful.

Theorem (3.14)

Let X be a random variable for which the MGF $M_X(t)$ exists. Then

$$\mu_r' = \mathrm{E}(X^r) = \left.\frac{\mathrm{d}^r M_X(t)}{\mathrm{d}t^r}\right|_{t=0}$$

Also useful: one may "invert" the MGF, $f(x) = \frac{1}{2\pi}\int_0^{2\pi} M_X(\mathrm{i}\theta)e^{-\mathrm{i}x\theta}\mathrm{d}\theta$.

Example

Consider the pdf

$$f(x) = e^{-x}\mathbb{I}_{(0,\infty)}(x) \qquad \text{(pdf of an exponential distribution).}$$

The MGF is given by

$$\begin{aligned} M_X(t) &= \int_{-\infty}^{\infty} e^{tx}e^{-x}\mathbb{I}_{(0,\infty)}(x)\mathrm{d}x = \int_0^{\infty} e^{x(t-1)}\mathrm{d}x \\ &= \left[\frac{e^{x(t-1)}}{t-1}\right]_{x=0}^{x=\infty}\Bigg|_{t<1} = 0 - \frac{1}{t-1} = \frac{1}{1-t}. \end{aligned}$$

The mean and the 2nd non-central moment are given by

$$\mu = \frac{\mathrm{d}M_X(t)}{\mathrm{d}t}\bigg|_{t=0} = \frac{1}{(1-t)^2}\bigg|_{t=0} = 1, \qquad \mu_2' = \frac{\mathrm{d}^2M_X(t)}{\mathrm{d}t^2}\bigg|_{t=0} = \frac{2}{(1-t)^3}\bigg|_{t=0} = 2.$$

Taylor around $t = 0$

The MGF $M_X(t) = \mathrm{E}\left(e^{tX}\right)$ can be written as a series expansion in terms of the moments of the pdf of X:

$$
\begin{aligned}
M_X(t) = \mathrm{E}\left(e^{tX}\right) &= \mathrm{E}\left(e^{0X} + \frac{1}{1!}[Xe^{0x}]t + \frac{1}{2!}[X^2e^{0x}]t^2 + \frac{1}{3!}[X^3e^{0x}]t^3 + \cdots\right) \\
&= 1 + \mu_1' t + \frac{1}{2!}\mu_2' t^2 + \frac{1}{3!}\mu_3' t^3 + \cdots \\
&= \sum_{i=0}^{\infty} \frac{1}{i!} t^i \mu_i'.
\end{aligned}
$$

Therfore: if the MGF exists, it characterizes an infinite set of moments.

The existence of all moments does not imply the existence of the MGF, though.

Useful properties

Let $X_1, ..., X_n$ be independent random variables having MGFs $M_{X_i}(t)$, $i = 1, \ldots, n$. Then we get

- for $Y = aX_i + b$ the MGF

$$M_Y(t) = \mathrm{E}\left(e^{(aX_i+b)t}\right) = e^{bt} M_{X_i}(at);$$

- for $Y = \sum_{i=1}^n X_i$ the MGF

$$M_Y(t) = \mathrm{E}\left(e^{(\sum_{i=1}^n X_i)t}\right) = \underbrace{\mathrm{E}\left(\prod_{i=1}^n e^{X_i t}\right) = \prod_{i=1}^n \mathrm{E}\, e^{X_i t}}_{\text{by independence}} = \prod_{i=1}^n M_{X_i}(t);$$

- for $Y = \sum_{i=1}^n a_i X_i + b$ the MGF

$$M_Y(t) = e^{bt} \prod_{i=1}^n M_{X_i}(a_i t).$$

The second use

Theorem (3.15 (MGF Uniqueness Theorem; scalar case))

If an MGF exists for a random variable X having pdf $f(x)$, then

- *the MGF is unique;*
- *and, conversely, the MGF determines the pdf of X uniquely, at least up to a set of points having probability 0.*

This ...

- allows us e.g. to identify distributions:
  - if a distribution has the MGF $e^{t^2/2}$
  - (which, as we'll see in future classes, is the MGF of the standard normal)
  - then it must be the standard normal distribution, and
- holds for random vectors as well.

Example

Suppose Z has an MGF defined by $M_Z(t) = \frac{1}{1-t}$ for $|t| < 1$.

Now, consider the pdf

$$f(x) = e^{-x}\mathbb{I}_{(0,\infty)}(x), \quad \text{which has an MGF} \quad M_X(t) = \frac{1}{1-t}$$

(see the previous Example). Then, by the uniqueness theorem, the pdf of Z must be

$$Z \sim f(z) = e^{-z}\mathbb{I}_{(0,\infty)}(z).$$

(Almost everywhere.)

Outline

Functionals

Moments are functions of the cdf F, say $\tau(F)$. Since the argument of τ is a function, we call τ a functional.

Moments are so-called linear functionals (of the form $\int g(x)\mathrm{d}F(x)$); and they are not the only functionals of interest ...

Definition (Median)

Any number, b, satisfying

$$\mathrm{P}\left(X \leq b\right) \geq 1/2 \qquad \text{and} \qquad \mathrm{P}\left(X \geq b\right) \geq 1/2$$

is called a **median** of X and is denoted by med(X).

The median is an alternative location measure for the distribution.[1]

[1] And statistical folklore says it's more robust to so-called outliers.

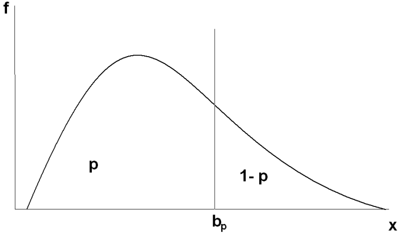
The median is a special quantile

Definition (Quantile)

A number q_p is a **quantile** of X of order p (or the $(100p)$th percentile of X) if

$$\mathrm{P}(X \leq q_p) \geq p \qquad \text{and} \qquad \mathrm{P}(X \geq q_p) \geq 1 - p.$$

In case of non-uniqueness, a convenient choice is $q_p = \inf\{x : F(x) \geq p\}$.



And unique quantiles have special properties

- Quantiles at any p are unique if the cdf F is continuous; $q_p = F^{-1}(p)$.
- Quantiles are (location) equivariant, i.e. if X has quantile q_p, then $X + \mu$ has quantile $q_p + \mu$
- Quantiles are linear in scale, i.e. if X has quantile q_p, then σX has quantile σq_p
- Unique quantiles minimize a certain expectation,

$$q_p = \arg\min_{q^*} \mathrm{E}\left(\rho_p\left(X - q^*\right)\right)$$

where ρ_p is the so-called quantile check function,

$$\rho_p(u) = \begin{cases} -(1-p)u, & u < 0; \\ pu, & u \geq 0. \end{cases}$$

Outline

Coming up

Joint and conditional moments