CSE 312

Foundations of Computing II

Lecture 22: Moments

PAUL G. ALLEN SCHOOL
OF COMPUTER SCIENCE & ENGINEERING

Stefano Tessaro

tessaro@cs.washington.edu

Things we mentioned, but did not prove:

- If $X \sim \mathcal{N}(\mu, \sigma^2)$, then $aX + b \sim \mathcal{N}(a\mu + b, a^2\sigma^2)$.
- If $X_1 \sim \mathcal{N}(\mu_1, \sigma_1^2)$ and $X_2 \sim \mathcal{N}(\mu_2, \sigma_2^2)$, then $X_1 + X_2 \sim \mathcal{N}(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$.
- The Central Limit Theorem (CLT).
  (Aka. "Everything" converges to a Gaussian!)

Reminder

$$e^x = \sum_{i=0}^{\infty} \frac{x^i}{i!}$$

We are going to use this <u>many</u> times today!

Moments

Definition. The k-th **moment** of a random variable X is $\mathbb{E}(X^k)$.

1st moment = expectation $\mathbb{E}(X)$

1st moment and 2nd moment → variance $\mathrm{Var}(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2$

Generally, a random variable is determined uniquely by its moments.

… let's make this more formal!

Moment Generating Functions

Definition. The **moment generating function (MGF)** of X is the function $M_X : \mathbb{R} \to \mathbb{R}$

$$M_X(t) = \mathbb{E}(e^{tX}).$$

$$M_X(t) = \mathbb{E}(e^{tX}) = \mathbb{E}\left(\sum_{i=0}^{\infty} \frac{(tX)^i}{i!}\right)$$

$$= 1 + \mathbb{E}(X)t + \mathbb{E}(X^2)\frac{t^2}{2} + \mathbb{E}(X^3)\frac{t^3}{6} + \cdots$$

MGFs – Basic Properties

Theorem. X and Y are identically distributed if and only if $M_X = M_Y$.

Theorem. If X and Y are independent, then for all $t \in \mathbb{R}$,

$$M_{X+Y}(t) = M_X(t) \cdot M_Y(t)$$

Proof.

$$M_{X+Y}(t) = \mathbb{E}\left(e^{t(X+Y)}\right) = \mathbb{E}(e^{tX}e^{tY}) = \mathbb{E}(e^{tX}) \cdot \mathbb{E}(e^{tY}) = M_X(t) \cdot M_Y(t)$$

Example – MGF of Poisson

Recall: $X \sim \text{Poi}(\lambda)$: $\mathbb{P}(X = i) = e^{-\lambda} \frac{\lambda^i}{i!}$

$$M_X(t) = \mathbb{E}(e^{tX}) = \sum_{i=0}^{\infty} \mathbb{P}(X = i) \cdot e^{ti}$$

$$= \sum_{i=0}^{\infty} e^{-\lambda} \frac{\lambda^i}{i!} \cdot e^{ti} = e^{-\lambda} \sum_{i=0}^{\infty} \frac{(e^t \lambda)^i}{i!}$$

$$= e^{-\lambda} e^{\lambda e^t} = e^{\lambda(e^t - 1)}$$

Example – Sum of Poissons

Claim. If $X_1 \sim \mathrm{Poi}(\lambda_1)$ and $X_2 \sim \mathrm{Poi}(\lambda_2)$ are independent, then

$$X_1 + X_2 \sim \mathrm{Poi}(\lambda_1 + \lambda_2)$$

Proof.

Previous slide

$$M_{X_1}(t) = e^{\lambda_1(e^t-1)}$$

$$M_{X_2}(t) = e^{\lambda_2(e^t-1)}$$

Reminder: If X and Y are independent, then $M_{X+Y}(t) = M_X(t) \cdot M_Y(t)$

$$M_{X_1+X_2}(t) = M_{X_1}(t) \cdot M_{X_2}(t) = e^{\lambda_1(e^t-1)} \cdot e^{\lambda_2(e^t-1)} = \underline{\mathrm{e}^{(\lambda_1+\lambda_2)(e^t-1)}}$$

$X_1 + X_2 \sim \mathrm{Poi}(\lambda_1 + \lambda_2)$

MGF of the Normal Distribution

Theorem. If $X \sim \mathcal{N}(\mu, \sigma^2)$, then $M_X(t) = e^{t\mu + \frac{t^2\sigma^2}{2}}$

We will prove it below, but first, some interesting consequences!

MGF of Normal – Applications

Theorem. If $X \sim \mathcal{N}(\mu, \sigma^2)$, then $M_X(t) = e^{t\mu + \frac{t^2\sigma^2}{2}}$

Fact 1. If $X \sim \mathcal{N}(\mu, \sigma^2)$, then $Y = aX + b \sim \mathcal{N}(a\mu + b, a^2\sigma^2)$

$$M_X(t) = e^{t\mu + \frac{t^2\sigma^2}{2}}$$

$$M_Y(t) = \mathbb{E}\left(e^{t(aX+b)}\right) = \mathbb{E}\left(e^{taX} e^{tb}\right)$$

$$= e^{tb}\mathbb{E}(e^{(ta)X})$$

$$= e^{tb} M_X(ta)$$

$$= e^{tb} e^{ta\mu + \frac{t^2a^2\sigma^2}{2}} = e^{t(a\mu+b) + \frac{t^2a^2\sigma^2}{2}}$$

← MGF of $\mathcal{N}(a\mu + b, a^2\sigma^2)$

MGF of Normal – Applications

Theorem. If $X \sim \mathcal{N}(\mu, \sigma^2)$, then $M_X(t) = e^{t\mu + \frac{t^2\sigma^2}{2}}$

Fact 2. If $X_1, \dots, X_n$ independent and $X_i \sim \mathcal{N}(\mu_i, \sigma_i^2)$, then

$$X_1 + \cdots + X_n \sim \mathcal{N}(\mu_1 + \cdots + \mu_n, \sigma_1^2 + \cdots + \sigma_n^2)$$

$$M_{X_1+\cdots+X_n}(t) = M_{X_1}(t) \cdots M_{X_n}(t)$$

$$= e^{t\mu_1 + \frac{t^2\sigma_1^2}{2}} \cdots e^{t\mu_n + \frac{t^2\sigma_n^2}{2}} = e^{t(\mu_1 + \cdots + \mu_n) + \frac{t^2(\sigma_1^2 + \cdots + \sigma_n^2)}{2}}$$

MGF of $\mathcal{N}(\mu_1 + \cdots + \mu_n, \sigma_1^2 + \cdots + \sigma_n^2)$

Try a direct proof for both facts?

MGF of the Normal Distribution – Proof – Standard Normal

Recall: If $X \sim \mathcal{N}(0,1)$, then $f_X(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$

$$M_X(t) = \mathbb{E}(e^{tX}) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{tx} e^{-x^2/2} \mathrm{d}x = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{tx - x^2/2} \mathrm{d}x$$

$$tx - \frac{x^2}{2} = \frac{2tx - x^2}{2} = \frac{t^2 - (x-t)^2}{2}$$

$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{\frac{t^2}{2} - (x-t)^2/2} \mathrm{d}x = e^{t^2/2} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-(x-t)^2/2} \mathrm{d}x = e^{t^2/2}$$

MGF of the Normal Distribution – General Proof (1/2)

Recall: If $X \sim \mathcal{N}(\mu, \sigma^2)$, then $f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/2\sigma^2}$

$$M_X(t) = \mathbb{E}(e^{tX}) = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{+\infty} e^{tx} e^{-(x-\mu)^2/2\sigma^2} \mathrm{d}x$$

$$= \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{+\infty} e^{t(z\sigma+\mu)-z^2/2} \mathrm{d}x$$

$$= e^{t\mu} \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{+\infty} e^{tz\sigma-z^2/2} \mathrm{d}x$$

$$= e^{t\mu} \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{+\infty} e^{tz\sigma-z^2/2} \frac{\mathrm{d}x}{\mathrm{d}z} \mathrm{d}z$$

$$z = \frac{x-\mu}{\sigma}$$

MGF of the Normal Distribution – General Proof (2/2)

Recall: If $X \sim \mathcal{N}(\mu, \sigma^2)$, then $f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/2\sigma^2}$

$$M_X(t) = \mathbb{E}(e^{tX}) = e^{t\mu} \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{+\infty} e^{tz\sigma - z^2/2} \frac{\mathrm{d}x}{\mathrm{d}z} \mathrm{d}z$$

$$z = \frac{x-\mu}{\sigma} \qquad \frac{\mathrm{d}x}{\mathrm{d}z} = \sigma$$

Rewrite x as a function of z, and take derivative!

$$= e^{t\mu} \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{+\infty} e^{tz\sigma - z^2/2} \sigma \mathrm{d}z$$

$$= e^{t\mu} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{tz\sigma - z^2/2} \mathrm{d}z$$

$$= e^{t\mu} e^{\frac{t^2\sigma^2}{2}} = e^{t\mu + \frac{t^2\sigma^2}{2}}$$

Proof of the CLT

Theorem. (Central Limit Theorem) The CDF of Y_n converges to the CDF of the standard normal $\mathcal{N}(0,1)$, i.e.,

$$\lim_{n\to\infty} \mathbb{P}(Y_n \le y) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{y} e^{-x^2/2} \mathrm{d}x$$

$$Y_n = \frac{X_1 + \cdots + X_n - n\mu}{\sigma\sqrt{n}}$$

$X_1, \dots, X_n$ iid with mean μ and variance σ^2

Proof shows that $M_{Y_n}(t) \to e^{t^2/2}$ as $n \to \infty$

Let's do this for the case $\sigma = 1$ and $\mu = 0$.

$$Y_n = \frac{X_1 + \cdots + X_n}{\sqrt{n}}$$

$$M_{Y_n}(t) = \mathbb{E}(e^{tX}) = \mathbb{E}\left(e^{t(X_1 + \cdots + X_n)/\sqrt{n}}\right)$$

$$= \prod_{i=1}^{n} \mathbb{E}\left(e^{tX_i/\sqrt{n}}\right) = \left(\mathbb{E}\left(e^{tX/\sqrt{n}}\right)\right)^n$$

X has same distribution as $X_1, \dots, X_n$

Proof of the CLT

$$M_{Y_n}(t) = \left(\mathbb{E}(e^{tX/\sqrt{n}})\right)^n$$

$$\mathbb{E}(e^{tX/\sqrt{n}}) = 1 + \mathbb{E}(X)\frac{t}{\sqrt{n}} + \mathbb{E}(X^2)\frac{t^2}{2n} + \mathbb{E}(X^3)\frac{t^3}{6n^{1.5}} + \cdots$$

But note that: $\mathbb{E}(X) = 0 \quad 1 = \sigma^2 = \mathbb{E}(X^2) - \mathbb{E}(X)^2 = \mathbb{E}(X^2)$

$$\mathbb{E}(e^{tX/\sqrt{n}}) = 1 + \frac{t^2}{2n} + \mathbb{E}(X^3)\frac{t^3}{6n^{1.5}} + \mathbb{E}(X^4)\frac{t^2}{24n^2} + \cdots$$

$$= 1 + \frac{t^2}{2n}\left[1 + \frac{\mathbb{E}(X^3)t}{3n^{0.5}} + \frac{\mathbb{E}(X^4)t^2}{12n} + \cdots\right] \approx 1 + \frac{t^2}{2n}$$ as $n \to \infty$

$$M_{Y_n}(t) = \left(\mathbb{E}(e^{tX/\sqrt{n}})\right)^n \approx \left(1 + \frac{t^2}{2n}\right)^n \to e^{t^2/2}$$ as $n \to \infty$