Slide 1

Things you need to know from probability theory

(Not to be read by experts)

Chris Hughes

University of Michigan

Rochester 1 June 2006

Slide 2

Discrete events, e.g. roll of die

$$\mathbb{P}\{X = n\} = p_n$$

Continuous events, e.g. time spent queuing for lunch

$$\begin{aligned}\mathbb{P}\{Y \leq t\} &= \int_{\mathbb{R}} \mathbb{1}_{\{x \leq t\}} \mathrm{d}\mu(x) \\ &= \int_{-\infty}^{t} p(x)\, \mathrm{d}x\end{aligned}$$

assuming density exists. (Might contain point masses = atoms).

$p(x)$ is the density, and $p(x)\mathrm{d}x$ is $\mathbb{P}\{Y \in \mathrm{d}x\} = \mathbb{P}\{Y \in (x, x + \mathrm{d}x)\}$

- Non-negative
- Total mass 1.

Slide 3

- Binomial ($0 \leq n \leq N$)

$$p_n = \binom{N}{n} p^n (1-p)^{N-n}$$

- Normal ($x \in \mathbb{R}$)

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

- Gamma ($x > 0$)

$$p_{\alpha,\beta}(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}$$

- Cauchy ($x \in \mathbb{R}$)

$$p(x) = \frac{1}{\pi(x^2+1)}$$

Slide 4

Definition: A random variable is a measurable function from a probability space, $(\Omega, A, \mathbb{P})$ to a measurable space.

- Ω: Set of outcomes (eg the reals)
- A: Set of events (eg Borel measurable subsets of the reals)
- $\mathbb{P}$: Probability / measure

Often abbreviate $\mathbb{P}\{\omega \in \Omega \, : \, X(\omega) > 0\}$ to $\mathbb{P}\{X > 0\}$.

Formally: X is an r.v. if $\{\omega \in \Omega \, : \, X(\omega) \leq r\} \in A$ for all real r.

E.g. $X = |Z_U(0)|$ on $(\mathbb{R}_{\geq 0}, A, \mu)$ where μ is Haar measure on $U(N)$.

So $\mathbb{P}\{|Z_U(0)| \geq 101\} = \mu\{|Z_U(0)| \geq 101\}$ is the measure of $N \times N$ unitary matrices such that $|Z_U(0)| \geq 101$.

Slide 5

Expectation denotes taking an average.

$$\mathbb{E}[f(X)] = \sum_{n=-\infty}^{\infty} f(n)p_n$$

or (with density)

$$\mathbb{E}[f(X)] = \int_{-\infty}^{\infty} f(x)p(x)\mathrm{d}x$$

or (with matrices)

$$\mathbb{E}[f(U)] = \int_{U(N)} f(U)\mathrm{d}\mu(U)$$

Also written as $\langle f(X) \rangle$.

Sometimes the space one is averaging against is indicated, eg. $\mathbb{E}_{U(N)}[|Z_U(0)|^2]$ or $\langle |Z_U(0)|^2 \rangle_{U(N)} = \int |Z_U(0)|^2 \mathrm{d}\mu(U)$.

Slide 6

- Mean (average value): $f(x) = x$.

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} xp(x)\ \mathrm{d}x \quad (= \mu \text{ say})$$

- Variance (measure of how close to mean): $f(x) = (x-\mu)^2$

$$\mathbb{E}[(X-\mu)^2] = \int_{-\infty}^{\infty} (x-\mu)^2 p(x)\mathrm{d}x \quad (= \sigma^2 \text{ say})$$

  Note

$$\begin{aligned}\mathbb{E}[(X-\mu)^2] &= \mathbb{E}[X^2] - 2\mathbb{E}[\mu X] + \mathbb{E}[\mu^2] = \mathbb{E}[X^2] - 2\mu\mathbb{E}[X] + \mu^2 \\ &= \mathbb{E}[X^2] - \mu^2\end{aligned}$$

- kth moment: $f(x) = x^k$

$$\mathbb{E}[X^k] = \int_{-\infty}^{\infty} x^k p(x)\ \mathrm{d}x \quad (= m_k \text{ say})$$

Slide 7

- Moment generating function: $f(x) = e^{\lambda x}$

$$\begin{aligned} \mathbb{E}[e^{\lambda X}] &= \int_{-\infty}^{\infty} e^{\lambda x} p(x)\, \mathrm{d}x \\ &= \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} m_k \end{aligned}$$

  It is legal to switch the sum and the integral if the moments m_k are not too big.

- Characteristic function: $f(x) = e^{\mathrm{i}\lambda x}$

$$\mathbb{E}[e^{\mathrm{i}\lambda X}] = \int_{-\infty}^{\infty} e^{\mathrm{i}\lambda x} p(x)\, \mathrm{d}x$$

  Note the characteristic function always exists for real λ.

Slide 8

- Cumulants: $f(x) =$??? [There's no simple function]

$$\log \mathbb{E}[e^{\lambda X}] = \sum_{n=1}^{\infty} C_n \frac{\lambda^n}{n!}$$

  Identify the λ^n term in

$$\sum_{n=1}^{\infty} C_n \frac{\lambda^n}{n!} = \log\left(\sum_{k=0}^{\infty} \frac{\lambda^k}{k!} m_k\right) = \log\left(1 + \sum_{k=1}^{\infty} \frac{\lambda^k}{k!} m_k\right)$$

  $C_0 = 0$
  $C_1 = m_1 = \mathbb{E}[X] = \mu$
  $C_2 = m_2 - m_1^2 = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = \sigma^2$
  $C_3 = m_3 - 3m_1m_2 + 2m_1^3$
  etc.

Slide 9

Example. Normal random variable.

$$p(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

Moment generating function

$$\int_{-\infty}^{\infty} e^{\lambda x} \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) \mathrm{d}x = \exp\left(\lambda\mu + \tfrac{1}{2}\lambda^2\sigma^2\right)$$

Cumulants: $C_1 = \mu$ and $C_2 = \sigma^2$ and $C_n = 0$ for $n \geq 3$.

Centered Moments: If $\mu = 0$,

$$\begin{aligned} m_{2k} &= \frac{(2k)!}{2^k k!}\sigma^{2k} \\ m_{2k+1} &= 0 \end{aligned}$$

The first few moments are $1, 3, 15, 105, 945, \ldots$.

Slide 10

Random matrix example:

Moment generating function of $\log|Z_U(0)|$

$$\begin{aligned} \mathbb{E}\left[e^{\lambda \log|Z_U(0)|}\right] &= \int e^{\lambda \log|Z_U(0)|} \mathrm{d}\mu \\ &= \prod_{j=1}^{N} \frac{\Gamma(j)\Gamma(j+\lambda)}{\Gamma(j+\lambda/2)^2} \\ &= M_N(\lambda) \end{aligned}$$

But this is also the moments of $|Z_U(0)|$ (and not just integer moments)

$$\mathbb{E}\left[|Z_U(0)|^\lambda\right] = \int |Z_U(0)|^\lambda \, \mathrm{d}\mu$$

Slide 11

Let $p(x)$ be the value distribution of $|Z_U(0)|$.

These integrals can be rewritten:

$$M_N(\lambda) = \mathbb{E}\left[|Z_U(0)|^\lambda\right] = \int_0^\infty x^\lambda p(x)\,\mathrm{d}x$$

The function $p(x)$ can be recovered by Mellin inversion:

If $M_N(\lambda)$ is analytic in the strip $a < \mathfrak{Re}(\lambda) < b$ and if $|M_N(\lambda)| \ll |\lambda|^{-2}$ then $p(x)$ exists and is continuous and is given by

$$p(x) = \frac{1}{2\pi \mathrm{i} x} \int_{(c)} x^{-\lambda} M_N(\lambda)\,\mathrm{d}\lambda$$

for any $a < c < b - 1$. The integral is taken up the vertical line with $\mathfrak{Re}(\lambda) = c$.

Slide 12

So if I know the complex moments, and they are sufficiency nice, then I can deduce the density function via Mellin inversion.

Equivalently: Moment generating function (or characteristic function) is sufficiently nice (exists in any open set containing the origin), then get density by Fourier inversion.

Equivalently: If integer moments are sufficiently nice, then get mgf, so get density.

Slide 13

If m_k grow sufficiently slowly so that

$$\sum_{n=1}^{\infty} \frac{\lambda^{2n}}{(2n)!} m_{2n}$$

converges then the distribution is uniquely determined from its moments.

Better: (Carleman) A distribution on the reals is uniquely determined by its moments if

$$\sum_{n=1}^{\infty} \frac{1}{m_{2n}^{1/(2n)}} = \infty$$

In particular the Gaussian distribution is uniquely determined by its integer moments.

Slide 14

The log-normal is *not* determined by its moments (Steiltjes)

X has a lognormal distribution if $\log X$ has a normal distribution.

Probability density function (when $\log X$ has mean zero, variance 1)

$$p(x) = \frac{1}{x\sigma\sqrt{2\pi}} \exp\left(-(\log x)^2/2\right)$$

for $x > 0$.

Moments: $m_k = e^{k^2/2}$

Slide 15

Consider

$$p_\epsilon(x) = p(x)(1 + \epsilon \sin(2\pi \log x))$$

for $0 < \epsilon < 1$.

This is another (different) probability density with the same (integer) moments. Note $p_\epsilon(x) > 0$.

It suffices to show

$$\int_0^\infty x^k p(x) \sin(2\pi \log x)\,\mathrm{d}x = 0 \quad , \quad k = 0, 1, 2, \ldots$$

Change variables: $\log x = y + k$. Use $\sin(2\pi(y+k)) = \sin(2\pi y)$.

$$\frac{1}{\sqrt{2\pi}} e^{k^2/2} \int_{-\infty}^\infty e^{-y^2/2} \sin(2\pi y)\,\mathrm{d}y = 0$$

since the integrand is odd.

Slide 16

Sequence of random variables, X_n, $n = 1, 2, 3, \ldots$.

When can we say X_n converges to a random variable X? What does it mean?

Definition. Convergence in distribution / law: $X_n \xrightarrow{d} X$ if $\mathbb{P}\{X_n \le x\} \to \mathbb{P}\{X \le x\}$ for all x at which $\mathbb{P}\{X \le x\}$ is continuous.

Note that X does not necessarily live in the same probability space as X_n.

Definition. Weak convergence: $X_n \Rightarrow X$ if $\mathbb{E}[f(X_n)] \to \mathbb{E}[f(X)]$ for all bounded continuous functions.

Theorem. $X_n \xrightarrow{d} X$ iff $X_n \Rightarrow X$.

Slide 17

Theorem. If X is uniquely determined by its moments and

$$\mathbb{E}[(X_n)^k] \to \mathbb{E}[X^k]$$

for $k = 0, 1, 2, \ldots$, then $X_n \Rightarrow X$.

Example. Selberg showed that for any k,

$$\int_0^T \left(\frac{\log |\zeta(\frac{1}{2} + \mathrm{i}t)|}{\sqrt{\frac{1}{2} \log \log T}} \right)^{2k} \mathrm{d}t \to \frac{(2k)!}{2^k k!}$$

while the odd moments tend to zero, and hence we may conclude that

$$\frac{\log |\zeta(\frac{1}{2} + \mathrm{i}t)|}{\sqrt{\frac{1}{2} \log \log T}}$$

weakly converges to a Gaussian distribution with mean zero and variance 1.

Slide 18

If we know $X_n \Rightarrow X$ can we conclude $\mathbb{E}[(X_n)^k] \to \mathbb{E}[X^k]$?

No!

"Gliding humps argument"

Take $p(x)$ and let $p_n(x) = (1 - \frac{1}{n})p(x) + \frac{1}{n}\delta_{\{x=n^2\}}$.

Note that $p_n(x)$ is a density, but the mean does not converge, since

$$\int_{-\infty}^{\infty} x p_n(x)\, \mathrm{d}x = \left(1 - \frac{1}{n}\right)\mu + n$$

Slide 19

Definition. Convergence in probability. $X_n \xrightarrow{p} X$ if $\mathbb{P}\{|X_n - X| > \epsilon\} \to 0$.

Theorem. $X_n \xrightarrow{p} X$ implies $X_n \xrightarrow{d} X$.

Definition. Convergence in L_p. $X_n \to X$ in the pth mean if $\mathbb{E}[|X_n - X|^p] \to 0$.

Theorem. Convergence in pth mean implies convergence in probability. (Proof: Chebyshev's / Markov's inequality).

Slide 20

Central limit theorem:

If $X_1, X_2, \dots$ are iid rv's with mean μ and variance σ^2, and let $S_N = \sum_{n=1}^N X_n$. Then

$$\frac{S_N - N\mu}{\sigma\sqrt{N}} \Rightarrow \mathcal{N}(0,1)$$

Pseudo-proof: Look at mgf:

$$\begin{aligned}\mathbb{E}\left[\exp\left(\lambda\sum_{j=1}^{N}\left(\frac{X_j-\mu}{\sigma\sqrt{N}}\right)\right)\right] &= \left(\mathbb{E}\left[\exp\left(\lambda\frac{X_1-\mu}{\sigma\sqrt{N}}\right)\right]\right)^N \\ &= \left(1+\frac{\lambda^2}{2N}+O\left(\frac{\lambda^3}{N^{3/2}}\right)\right)^N \\ &\to e^{\lambda^2/2}\end{aligned}$$