Introduction to probability cheat-sheet

(version 1: 17.04.2024)

1. Classical probability

- Variations with repetitions: n^k
- Variations without repetition: $\frac{n!}{(n-k)!}$
- Permutations: $n!$
- Combinations: $\binom{n}{k}$

2. Axiomatic def. of probability

- Definition of σ-algebra $\mathcal{F} \subseteq 2^{\Omega}$:

  1. $\Omega \in \mathcal{F}$
  2. If $A \in \mathcal{F}$ then $A' \in \mathcal{F}$
  3. If $A_1, A_2, \ldots \in \mathcal{F}$ then $A_1 \cup A_2 \cup \ldots \in \mathcal{F}$

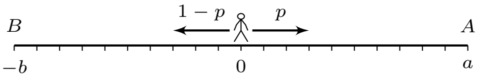
- Properties of σ-algebra: $\emptyset \in \mathcal{F}$; if $A, B \in \mathcal{F}$, then $A \cap B \in \mathcal{F}$, $A \setminus B \in \mathcal{F}$
- Properties of probability:

  - $P(\emptyset) = 0$, $P(A') = 1 - P(A)$
  - If $A \subseteq B$ then $P(B \setminus A) = P(B) - P(A)$
  - $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
  - $P(A_1 \cup \ldots \cup A_n) \leqslant P(A_1) + \ldots + P(A_n)$, equality for disjoint events ($A_i \cap A_j = \emptyset$ for $i \neq j$)

3. Conditional probability

- Definition: $P(A|B) = \frac{P(A \cap B)}{P(B)}$ for $P(B) > 0$
- $P(A \cap B) = P(A|B)P(B) = P(B|A)P(A)$
- Chain rule:
  $P(A_1 \cap \ldots \cap A_n) = P(A_1)P(A_2|A_1) \cdot \ldots \cdot P(A_n|A_1 \cap \ldots \cap A_{n-1})$
- Partition $A_1, \ldots, A_n$: $A_i \cap A_j = \emptyset$ for $i \neq j$, and $A_1 \cup \ldots \cup A_n = \Omega$
- Total probability: if $A_1, \ldots, A_n$ – partition:
  $P(B) = \sum_{i=1}^{n} P(A_i)P(B|A_i)$
- Bayes' rule: if $A_1, \ldots, A_n$ – partition:
  $P(A_i|B) = \frac{P(B|A_i)P(A_i)}{P(B)} = \frac{P(B|A_i)P(A_i)}{\sum_{j=1}^{n} P(B|A_j)P(A_j)}$

4. Independence

- Definition $P(A \cap B) = P(A)P(B)$.
  More generally: $A_1, \ldots, A_n$ independent if for each $S \subseteq \{1, 2, \ldots, n\}$: $P\left(\bigcap_{i \in S} A_i\right) = \prod_{i \in S} P(A_i)$
- If $A \perp B$ then $A \perp B'$, $A' \perp B$, $A' \perp B'$
- If $A_1, \ldots, A_n$ – independent then
  $P(A_1 \cup \ldots \cup A_n) = 1 - P(A'_1) \cdot \ldots \cdot P(A'_n)$
- Conditional independence (given C):
  $P(A \cap B|C) = P(A|C)P(B|C)$
- Random walk: 

  - Prob. of reaching A:
  $$P(A) = \begin{cases} \frac{b}{a+b} & (p = \frac{1}{2}) \\ \frac{\left(\frac{p}{1-p}\right)^a - \left(\frac{p}{1-p}\right)^{a+b}}{1 - \left(\frac{p}{1-p}\right)^{a+b}} & (p \neq \frac{1}{2}) \end{cases}$$
  - Prob. of reaching B: $P(B) = 1 - P(A)$

5. Random variables

- Definition: measurable function $X \colon \Omega \to \mathbb{R}$
- Distribution of random variable: measure P_X over $\mathbb{R}$ with Borel σ-algebra, such that $P_X(A) = P(X \in A) = P(X^{-1}(A))$
- C.d.f.: $F_X(x) = P(X \leqslant x)$
- Properties of F_X: nondecreasing; $F(\infty) = 1$, $F(-\infty) = 0$; $P(a < X \leqslant b) = F(b) - F(a)$
- Degenerate distribution: $P(X = c) = 1$
- Uniform distribution: $X \in \{x_1, \ldots, x_n\}$, $P(X = x_i) = \frac{1}{n}$
- Bernoulli distribution $B(p)$: $X \in \{0, 1\}$, $P(X = 1) = p$, $P(X = 0) = 1 - p$
- Binomial distribution $B(n, p)$: $X \in \{0, 1, \ldots, n\}$, $P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$
- Geometric distribution $G_1(p)$: $X \in \{1, 2, \ldots\}$, $P(X = k) = (1-p)^{k-1} p$
- Geometric distribution $G_0(p)$: $X \in \{0, 1, \ldots\}$, $P(X = k) = (1-p)^k p$
- For $X \sim G_1(p)$: $P(X > k) = (1-p)^k$
- Memorylessness $X \sim G_1(p)$: $P(X > k + \ell | X > k) = P(X > \ell)$
- Negative binomial distribution $NB(r, p)$: $P(X = k) = \binom{r+k-1}{r-1} (1-p)^r p^k$
- Poisson distribution $\text{Pois}(\lambda)$: $X \in \{0, 1, \ldots\}$, $P(X = k) = \frac{\lambda^k}{k!} e^{-\lambda}$
- $B(n, p) \to \text{Pois}(\lambda)$ for $n \to \infty$ and $\lambda = np$

6. Moments of random variables

- For $X \in \{0, 1, \ldots\}$: $EX = \sum_{k=1}^{\infty} P(X \geqslant k)$
- For $Y = f(X)$: $EY = \sum_x f(x) P(X = x)$
- Linearity: $E(aX + b) = aEX + b$
- $D^2(X) = E\big((X - EX)^2\big) = E(X^2) - (EX)^2$
- $D^2(aX + b) = a^2 D^2(X)$
- $D^2(X) \geqslant 0$ and $D^2(X) = 0 \iff X$ has degenerate distr.
- Expected value and variance

Distribution of X	EX	$D^2(X)$
$B(p)$	p	$p(1-p)$
$B(n,p)$	np	$np(1-p)$
$G_1(p)$	$\frac{1}{p}$	$\frac{1-p}{p^2}$
$NB(r,p)$	$\frac{rp}{1-p}$	$\frac{rp}{(1-p)^2}$
$\text{Pois}(\lambda)$	λ	λ

- k-th order moment: $m_k = E(X^k)$
- k-th order central moment: $\mu_k = E\big((X - EX)^k\big)$
- Markov's inequality: for nonnegative X and $a > 0$: $P(X \geqslant a) \leqslant \frac{EX}{a}$
- Chebyshev's inequality: $P(|X - EX| \geqslant \epsilon) \leqslant \frac{D^2(X)}{\epsilon^2}$
- For $X \sim B(n, p)$ the most probable value is: (a) $\lfloor (n+1)p \rfloor$ if $(n+1)p$ is non-integer; (b) $(n+1)p$ and $(n+1)p - 1$ (two values) if $(n+1)p$ is integer

7. Multidimensional random variables

- Marginal distribution: $P(X = x) = \sum_y P(X = x, Y = y)$
- Conditional distribution: $P(X = x | Y = y) = \frac{P(X = x, Y = y)}{P(Y = y)}$
- $P(X \in A) = \sum_y P(X \in A | Y = y) P(Y = y)$ (total prob.)
- Conditional expectation:
  $E(X|Y = y) = \sum_x x\, P(X = x | Y = y)$
- $E\big(E(X|Y)\big) = EX$ (tower rule)

8. Multidimensional random variables II

- $E(X_1 + \ldots + X_n) = EX_1 + \ldots + EX_n$
- $C(X,Y) = E\big((X-EX)(Y-EY)\big) = E(XY) - (EX)(EY)$
- $D^2(X \pm Y) = D^2(X) \pm 2C(X,Y) + D^2(Y)$
- $|C(X,Y)| \leqslant D(X)D(Y)$
- $\rho(X,Y) = \frac{C(X,Y)}{D(X)D(Y)} \in [-1,1]$
- Independence:
  $P(X_1 \in A_1, \ldots, X_n \in A_n) = P(X_1 \in A_1) \cdot \ldots \cdot P(X_n \in A_n)$
- $X_1, \ldots, X_n$ independent: $E(X_1 \cdot \ldots \cdot X_n) = EX_1 \cdot \ldots \cdot EX_n$
- X, Y independent: $C(X,Y) = 0$
- $X_1, \ldots, X_n$ independent:
  $D^2(X_1 \pm \ldots \pm X_n) = D^2(X_1) + \ldots + D^2(X_n)$
- If $X_1, \ldots, X_n \sim B(p)$ independent then
  $Y = \sum_{i=1}^n X_i \sim B(n,p)$

9. Continuous random variables

- For $Y = g(X)$ g differentiable and invertible:
  $f_Y(y) = f_X\big(h(y)\big)|h'(y)|$, where $h = g^{-1}$
- If $Y = g(x)$ then $EY = \int_{-\infty}^{\infty} g(x) f(x)\,\mathrm{d}x$
- Uniform distr. $\mathrm{Unif}[a,b]$: $f(x) = \frac{1}{b-a}$ for $x \in [a,b]$
  $E(X) = \frac{a+b}{2}$, $D^2(X) = \frac{(b-a)^2}{12}$
- Exponential distr. $\mathrm{Exp}(\lambda)$: $f(x) = \lambda e^{-\lambda x}$ for $x \geqslant 0$,
  $F(x) = 1 - e^{-\lambda x}$, $EX = \frac{1}{\lambda}$, $D^2(X) = \frac{1}{\lambda^2}$
- Memorylessness: if $X \sim \mathrm{Exp}(\lambda)$ then
  $P(X \geqslant b | X \geqslant a) = P(X \geqslant b - a)$
- Normal distribution $N(\mu, \sigma^2)$: $f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\big\{-\frac{(x-\mu)^2}{2\sigma^2}\big\}$,
  $EX = \mu$, $D^2(X) = \sigma^2$
- If $X \sim N(\mu, \sigma^2)$ then $aX + b \sim N(\mu a + b, a^2\sigma^2)$
- If $X \sim N(\mu, \sigma^2)$ then $Z = \frac{X-\mu}{\sigma} \sim N(0,1)$
- If $Z \sim N(0,1)$ then $X = \sigma Z + \mu \sim N(\mu, \sigma^2)$
- C.d.f. of $Z \sim N(0,1)$: $\Phi(z) = P(Z \leqslant z)$, $\Phi(-z) = 1 - \Phi(z)$

10. Continuous random variables II

- $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\,\mathrm{d}x\,\mathrm{d}y = 1$
- Marginal density: $f_X(x) = \int_{-\infty}^{\infty} f(x,y)\,\mathrm{d}y$
- Conditional density: $f_{Y|X}(y|x) = \frac{f(x,y)}{f(x)}$
- $f_Y(y) = \int_{-\infty}^{\infty} f_{Y|X}(y|x) f_X(x)\,\mathrm{d}x$
- Independent random variables:
  $f(x_1, \ldots, x_n) = f_{X_1}(x_1) \cdot \ldots \cdot f_{X_n}(x_n)$
- $X_1, \ldots, X_n$ – independent with c.d.f. F_X, $Y = \max_i\{X_i\}$,
  $Z = \min_i\{X_i\}$ then $F_Y(y) = F_X(y)^n$,
  $F_Z(z) = 1 - (1 - F_X(z))^n$
- If X, Y independent and $Z = X + Y$ then:
  $f_Z(z) = \int_{-\infty}^{\infty} f_X(x) f_Y(z-x)\,\mathrm{d}x$ (convolution)
- If $X \sim N(\mu_X, \sigma_X^2)$, $Y \sim N(\mu_Y, \sigma_Y^2)$, and X, Y – independent
  then: $Z = X + Y \sim N(\mu_X + \mu_Y, \sigma_X^2 + \sigma_Y^2)$
- $X_i \sim N(\mu_i, \sigma_i^2)$ – independent, $Z = \sum_{i=1}^n a_i X_i$, then:
  $Z \sim N\big(\sum_{i=1}^n a_i\mu_i, \sum_{i=1}^n a_i^2\sigma_i^2\big)$
- Z has distribution $\chi^2(k)$ if $Z = \sum_{i=1}^k X_i^2$ where $X_i \sim N(0,1)$,
  independent. $EZ = k$
- T has t-Student distribution, $t(k)$, if $T = \frac{X}{\sqrt{Z}}\sqrt{k}$, where
  $X \sim N(0,1)$, $Z \sim \chi^2(k)$, X and Z independent

11. Limit theorems I

- If $X_1, \ldots, X_n$ independent with the same distr., $EX_i = \mu$ and
  $D^2(X_i) = \sigma^2$ then $E\overline{X}_n = \mu$ and $D^2(\overline{X}_n) = \frac{\sigma^2}{n}$
- $X_n \overset{\text{w.pr. }1}{\to} X$: $P\big(\lim_{n\to\infty} X_n = X\big) = 1$
- $X_n \overset{P}{\to} X$: $\forall \epsilon > 0, \lim_{n\to\infty} P(|X_n - X| > \epsilon) = 0$
- $X_n \overset{\text{w.pr. }1}{\to} X \Rightarrow X_n \overset{P}{\to} X \Rightarrow X_n \overset{D}{\to} X$
- (Strong) Bernoulli LLN: if $X_1, \ldots, X_n \sim B(p)$ independent,
  then $\overline{X}_n \overset{\text{w.pr. }1}{\to} p$
- (Strong) Khinchin LLN: if $X_1, \ldots, X_n$ independent with the
  same distr., $EX = \mu$, $D^2(X) < \infty$ then $\overline{X}_n \overset{\text{w.pr. }1}{\to} \mu$

12. Limit theorems II

- For $U = \frac{X - EX}{D(X)}$: $EU = 0$, $D^2(U) = 1$
- $X_n \overset{D}{\to} X$: $\lim_{n\to\infty} F_{X_n}(x) = F_X(x)$ at every continuity point
  of F_X
- Moivre-Laplace theorem: if $X_1, \ldots, X_n \sim B(p)$ independent,
  then $U = \frac{S_n - np}{\sqrt{np(1-p)}} = \frac{\overline{X}_n - p}{\sqrt{p(1-p)}}\sqrt{n} \overset{D}{\to} Z \sim N(0,1)$
- Lindeberg-Levy theorem: if $X_1, \ldots, X_n$ independent with the
  same distr., $EX = \mu$, $D^2(X) = \sigma^2$ then:
  $U = \frac{\overline{X}_n - \mu}{\sigma}\sqrt{n} \overset{D}{\to} Z \sim N(0,1)$
- Conclusion: if $S_n \sim B(n,p)$ then S_n can be approximated by
  $X \sim N(np, np(1-p))$ (condition: $np \geqslant 5$ i $n(1-p) \geqslant 5$)
- Moment generating function: $M_X(t) = E(e^{tX})$

distr. X	$M_X(t)$	distr. X	$M_X(t)$
$B(p)$	$pe^t + 1 - p$	$B(n,p)$	$(pe^t + 1 - p)^n$
$\mathrm{Unif}[0,1]$	$\frac{e^t - 1}{t}$ (1 if $t = 0$)	$\mathrm{Exp}(\lambda)$	$\frac{\lambda}{\lambda - t}$
$\mathrm{Pois}(\lambda)$	$e^{\lambda(e^t - 1)}$	$N(\mu, \sigma^2)$	$e^{\mu t + \frac{1}{2}\sigma^2 t^2}$

- $M_X(0) = 1$, $M_X^{(k)}(0) = E(X^k)$, $M_{aX+b}(t) = e^{bt} M_X(at)$,
- $M_{X+Y}(t) = M_X(t) M_Y(t)$ for X, Y - independent