Probability review

• *Probability space* Ω - a finite (or for us at most countable) set endowed with a measure $p : \Omega \to \mathcal{R}$ satisfying:

$$\forall \omega \in \Omega;\ p(\omega) \geq 0$$

and

$$\sum_{\omega \in \Omega} p(\omega) = 1.$$

• *An event* $\mathbf{A} \subseteq \Omega$ - $\Pr[A] = \sum_{\omega \in A} p(\omega)$.

• *Random variable* $\mathbf{X}$ - $\mathbf{X} : \Omega \to \mathcal{R}$.

Example: If $\mathbf{X}$ is a random variable then for a fixed $t, t' \in \mathcal{R}$, $t \leq \mathbf{X} \leq t'$ and $\mathbf{X} > t$ are probabilistic events.

• *Two events* $\mathbf{A}$ *and* $\mathbf{B}$ *are independent* - $\Pr[\mathbf{A} \cap \mathbf{B}] = \Pr[\mathbf{A}] \cdot \Pr[\mathbf{B}]$.

• *Conditional probability of* $\mathbf{A}$ *given* $\mathbf{B}$ - $\Pr[\mathbf{A}|\mathbf{B}] = \Pr[\mathbf{A} \cap \mathbf{B}]/\Pr[\mathbf{B}]$.

Example: $\mathbf{A}$ and $\mathbf{B}$ are independent iff $\overline{\mathbf{A}}$ and $\mathbf{B}$ are independent iff ... iff $\Pr[\mathbf{A}|\mathbf{B}] = \Pr[\mathbf{A}]$.

• *For a random variable* $\mathbf{X}$ *and an event* $\mathbf{A}$, $\mathbf{X}$ *is independent of* $\mathbf{A}$ - for all $S \subseteq \mathcal{R}$, $\Pr[\mathbf{X} \in S|\mathbf{A}] = \Pr[\mathbf{X} \in S]$.

• *Two random variables* $\mathbf{X}$ *and* $\mathbf{Y}$ *are independent* - for all $S, T \subseteq \mathcal{R}$, $\mathbf{X} \in S$ and $\mathbf{Y} \in T$ are independent events.

• *Events* $\mathbf{A}_1, \mathbf{A}_2, \ldots, \mathbf{A}_n$ *are mutually independent* - for all $I \subseteq \{1, \ldots, n\}$,

$$\Pr[\bigcap_{i \in I} A_i \cap \bigcap_{i \notin I} \overline{A_i}] = \prod_{i \in I} \Pr[A_i] \cdot \prod_{i \notin I} \Pr[\overline{A_i}].$$

• *Random variables* $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n$ *are mutually independent* - for all $t_1, t_2, \ldots, t_n \in \mathcal{R}$, events $\mathbf{X}_1 = t_1$, $\mathbf{X}_2 = t_2$,...,$\mathbf{X}_n = t_n$ are mutually independent.

• *Expectation of a random variable* $\mathbf{X}$ - $\mathbf{E}[\mathbf{X}] = \sum_{\omega \in \Omega} p(\omega)\mathbf{X}(\omega)$.
Three easy claims:

Claim: (Linearity of expectation) For random variables $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n$

$$\mathbf{E}[\mathbf{X}_1 + \mathbf{X}_2 + \cdots \mathbf{X}_n] = \sum_{i=1}^{n} \mathbf{E}[\mathbf{X}_i].$$

Claim: For independent random variables $\mathbf{X}$ and $\mathbf{Y}$, $\mathbf{E}[\mathbf{X} \cdot \mathbf{Y}] = \mathbf{E}[\mathbf{X}] \cdot \mathbf{E}[\mathbf{Y}]$.

Claim: For a random variable $\mathbf{X} : \Omega \to \mathcal{N}$, $\mathbf{E}[\mathbf{X}] = \sum_{k=1}^{\infty} \mathbf{E}[\mathbf{X} \geq k]$.

Theorem: (Markov Inequality) For a non-negative random variable $\mathbf{X}$ and any $t \in \mathcal{R}$

$$\Pr[\mathbf{X} \geq t] \leq \frac{\mathbf{E}[\mathbf{X}]}{t}.$$

Proof: $\mathbf{E}[\mathbf{X}] = \sum_{\omega\in\Omega} p(\omega)\mathbf{X}(\omega) \geq \sum_{\omega\in\Omega,\ \mathbf{X}(\omega)\geq t} p(\omega)\mathbf{X}(\omega) \geq t\cdot\sum_{\omega\in\Omega,\ \mathbf{X}(\omega)\geq t} p(\omega) = t\cdot\Pr[\mathbf{X}\geq t]$. □

- *Variance* $\mathbf{Var}[\mathbf{X}]$ *of a random variable* $\mathbf{X}$ - $\mathbf{Var}[\mathbf{X}] = \mathbf{E}[(\mathbf{X}-\mu)^2]$ where $\mu = \mathbf{E}[\mathbf{X}]$.

Claim: For any random variable $\mathbf{X}$, $\mathbf{Var}[\mathbf{X}] = \mathbf{E}[\mathbf{X}^2] - (\mathbf{E}[\mathbf{X}])^2$.

Claim: For any random variable $\mathbf{X}$ and a constant c, $\mathbf{Var}[c\mathbf{X}] = c^2\mathbf{Var}[\mathbf{X}]$.

Claim: (Linearity of variance) For mutually independent random variables $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n$, $\mathbf{Var}[\mathbf{X}_1 + \mathbf{X}_2 + \cdots \mathbf{X}_n] = \mathbf{Var}[\mathbf{X}_1] + \mathbf{Var}[\mathbf{X}_2] + \cdots \mathbf{Var}[\mathbf{X}_n]$.

Theorem: (Chebyshev's inequality) Let $\mathbf{X}$ be a random variable. For any real number $a > 0$ it holds:

$$\Pr(|\mathbf{X} - \mathbf{E}[\mathbf{X}]| > a) \leq \frac{\mathbf{Var}[\mathbf{X}]}{a^2}.$$

Proof: Let $\mu = \mathbf{E}[\mathbf{X}]$. Consider the non-negative random variable $\mathbf{Y} = (\mathbf{X} - \mu)^2$. Clearly $\mathbf{E}[\mathbf{Y}] = \mathbf{Var}[\mathbf{X}]$. Using Markov inequality,

$$\begin{aligned}\Pr[|\mathbf{X} - \mu| > a] &= \Pr[\mathbf{Y} > a^2] \\ &\leq \frac{\mathbf{E}[\mathbf{Y}]}{a^2} \\ &= \frac{\mathbf{Var}[\mathbf{X}]}{a^2}.\end{aligned}$$

□

Theorem: (Chernoff Bounds) Let $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n$ be independent 0-1 random variables. Denote $p_i = \Pr[\mathbf{X}_i = 1]$, hence $1 - p_i = \Pr[\mathbf{X}_i = 0]$. Let $\mathbf{X} = \sum_{i=1}^n X_i$. Denote $\mu = \mathbf{E}[X] = \sum_{i=1}^n p_i$. For any $0 < \delta < 1$ it holds

$$\Pr[\mathbf{X} \geq (1+\delta)\mu] \leq \left[\frac{\mathrm{e}^{\delta}}{(1+\delta)^{(1+\delta)}}\right]^{\mu}$$

and

$$\Pr[\mathbf{X} \leq (1-\delta)\mu] \leq \mathrm{e}^{-\frac{1}{2}\mu\delta^2}.$$

Proof: For any real number $t > 0$,

$$\begin{aligned}\Pr[\mathbf{X} \geq (1+\delta)\mu] &= \Pr[t\mathbf{X} \geq t(1+\delta)\mu] \\ &= \Pr[\mathrm{e}^{t\mathbf{X}} \geq e^{t(1+\delta)\mu}]\end{aligned}$$

where based on $\mathbf{X}$ we define new random variables $t\mathbf{X}$ and $\mathrm{e}^{t\mathbf{X}}$. Notice, $\mathrm{e}^{t\mathbf{X}}$ is a non-negative random variable so one can apply the Markov inequality to obtain

$$\Pr[\mathrm{e}^{t\mathbf{X}} \geq e^{t(1+\delta)\mu}] \leq \frac{\mathbf{E}[e^{t\mathbf{X}}]}{e^{t(1+\delta)\mu}}.$$

Since all $\mathbf{X}_i$ are mutually independent, random variables $e^{t\mathbf{X}_i}$ are also mutually independent so

$$\mathbf{E}[\mathrm{e}^{t\mathbf{X}}] = \mathbf{E}[\mathrm{e}^{t\sum_i \mathbf{X}_i}] = \prod_{i=1}^{n} \mathbf{E}[\mathrm{e}^{t\mathbf{X}_i}].$$

We can evaluate $\mathbf{E}[\mathrm{e}^{t\mathbf{X}_i}]$

$$\mathbf{E}[\mathrm{e}^{t\mathbf{X}_i}] = p_i\mathrm{e}^t + (1-p_i)\cdot 1 = 1 + p_i(\mathrm{e}^t - 1) \le \mathrm{e}^{p_i(e^t-1)}.$$

where in the last step we have used $1 + x \le \mathrm{e}^x$ which holds for all x. (Look on the graph of functions $1 + x$ and e^x and their derivatives in $x = 0$.) Thus

$$\begin{aligned} \mathbf{E}[e^{tX}] &\le \prod_{i=1}^{n} \mathrm{e}^{p_i(e^t-1)} \\ &= \mathrm{e}^{\sum_{i=1}^{n} p_i(e^t-1)} \\ &= \mathrm{e}^{\mu(e^t-1)} \end{aligned}$$

By choosing $t = \ln(1+\delta)$ and rearanging terms we obtain

$$\begin{aligned} \Pr[\mathbf{X} \ge (1+\delta)\mu] &= \Pr[\mathrm{e}^{t\mathbf{X}} \ge e^{t(1+\delta)\mu}] \\ &\le \frac{\mathrm{e}^{\mu(e^t-1)}}{e^{t(1+\delta)\mu}} \\ &= \left[\frac{\mathrm{e}^{\delta}}{(1+\delta)^{(1+\delta)}}\right]^{\mu} \end{aligned}$$

That proofs the first bound. The second bound is obtained in a similar way:

$$\begin{aligned} \Pr[\mathbf{X} \le (1-\delta)\mu] &= \Pr[-t\mathbf{X} \ge -t(1-\delta)\mu] \\ &= \Pr[\mathrm{e}^{-t\mathbf{X}} \ge e^{-t(1-\delta)\mu}] \\ &\le \frac{\mathbf{E}[e^{-t\mathbf{X}}]}{e^{-t(1-\delta)\mu}}. \end{aligned}$$

Bounding $\mathbf{E}[e^{-t\mathbf{X}}]$ as before gives

$$\mathbf{E}[e^{-tX}] \le \mathrm{e}^{\mu(e^{-t}-1)}$$

By choosing $t = -\ln(1-\delta)$ and rearanging terms we obtain

$$\begin{aligned} \Pr[\mathbf{X} \le (1-\delta)\mu] &= \Pr[\mathrm{e}^{-t\mathbf{X}} \ge e^{-t(1-\delta)\mu}] \\ &\le \frac{\mathrm{e}^{\mu(e^{-t}-1)}}{e^{-t(1-\delta)\mu}} \\ &= \left[\frac{\mathrm{e}^{-\delta}}{(1-\delta)^{(1-\delta)}}\right]^{\mu} \end{aligned}$$

We use the well known expansion for $0 < \delta < 1$

$$\ln(1-\delta) = -\sum_{i=1}^{\infty} \frac{\delta^i}{i}$$

to obtain

$$\begin{aligned} (1-\delta)\ln(1-\delta) &= \sum_{i=1}^{\infty} \frac{\delta^{i+1}}{i} - \sum_{i=1}^{\infty} \frac{\delta^i}{i} \\ &= \sum_{i=2}^{\infty} \frac{\delta^i}{i(i-1)} - \delta \end{aligned}$$

Thus

$$(1-\delta)^{(1-\delta)} \quad \geq \quad \mathrm{e}^{\frac{\delta^2}{2}-\delta}$$

Hence

$$\Pr[\mathbf{X} \leq (1-\delta)\mu] \leq \mathrm{e}^{-\frac{\delta^2}{2}+\delta-\delta} = \mathrm{e}^{-\frac{\delta^2}{2}\mu}$$

□