

# PROBABILITY THEORY

MATH 154

## Unit 12: Chebyshev-Markov-Chernoff-Gibbs

### CHEBYSHEV

**12.1.** We use the short-hand notation  $P[X \geq c]$  for  $P[\{\omega \in \Omega \mid X(\omega) \geq c\}]$ .

**Theorem 1** (Chebyshev-Markov inequality). *Let  $h$  be a monotone function on  $\mathbb{R}$  with  $h \geq 0$ . For every  $c > 0$ , and  $h(X) \in \mathcal{L}^1$  we have*

$$h(c) \cdot P[X \geq c] \leq E[h(X)] .$$

*Proof.* Integrate the inequality  $h(c)1_{X \geq c} \leq h(X)$  and use the monotonicity and linearity of the expectation.  $\square$

**12.2.**  $h(x) = |x|$  leads to  $P[|X| \geq c] \leq \|X\|_1/c$  which implies for example the statement

$$E[|X|] = 0 \Rightarrow P[X = 0] = 1 .$$

**12.3.** For  $X \in \mathcal{L}^\infty$  we had used the moment generating function  $M_X(t) = E[e^{Xt}]$ . For every  $t \geq 0$  we have

$$e^{tc}P[X \geq c] \leq M_X(t) .$$

This gives the

**Theorem 2** (Chernoff bound).

$$P[X \geq c] \leq \inf_{t \geq 0} e^{-tc} M_X(t) .$$

**12.4.** An important case of the Chebyshev-Markov inequality is the **Chebyshev inequality**:

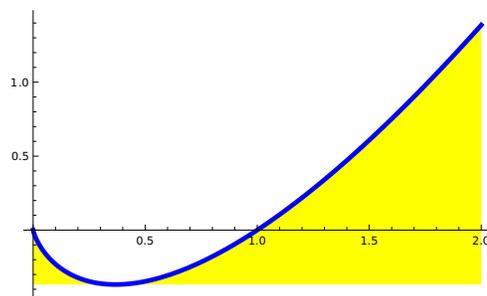
**Theorem 3** (Chebyshev inequality). *If  $X \in \mathcal{L}^2$ , then*

$$P[|X - E[X]| \geq c] \leq \frac{\text{Var}[X]}{c^2} .$$

*Proof.* Take  $h(x) = x^2$  and apply the Chebyshev-Markov inequality to the random variable  $Y = X - E[X]$  which is in  $\mathcal{L}^1$  because  $X \in \mathcal{L}^2$ .  $\square$

### ENTROPY

**12.5.** If  $(\Omega, \mathcal{A}, P)$  is a probability space and  $X \in \mathcal{S}$ ,  $X = \sum_{k=1}^n x_k 1_{A_k}$  is a random variable taking finitely many values  $x_k$ , its **entropy** is defined as  $\sum_k \log(\frac{1}{p_k})p_k$ , where  $p_k = P[X = x_k]$  and  $\log(\frac{1}{p_k})p_k = 0$  if  $p_k = 0$ . It has been introduced by Ludwig Boltzmann in statistical mechanics (" $S = k \log[W]$ ") and Claude Shannon in information theory.

FIGURE 1. The convex function  $x \log(x)$ .

**12.6.** The entropy of a finite sub-algebra  $\mathcal{B}$  of  $\mathcal{A}$  is defined as  $\sum_{A \in \mathcal{B}} P[A] \log(\frac{1}{P[A]})$ . This is well defined, if we understand that  $P[A] \log(\frac{1}{P[A]}) = 0$  for  $P[A] = 0$ . It is a common assumption to extend the convex function  $h(x) = x \log(x)$  to  $x = 0$  by setting it to  $h(0) = 0$ .

**12.7.** Any random variable  $X \in \mathcal{S}$  defines a finite probability distribution and so a finite set  $p_k$  of probabilities. We now take two distributions with  $p_k, q_k, k = 1, \dots, n$ . The **relative entropy** is defined as

$$D[p, q] = \sum_k p_k \log\left(\frac{p_k}{q_k}\right),$$

We can rewrite this as  $S(p, q) - S(p)$ , where

$$S(p, q) = \sum_k p_k \log\left(\frac{1}{q_k}\right)$$

is the **cross entropy**.

**12.8.** The relative entropy is also known as the **Kullback-Leibler divergence**. In nice cases, this goes over to more general random variables like continuous distributions where entropy is  $S(f) = \int f(x) \log(\frac{1}{f(x)}) dx$  and  $D[f, g] = \int f(x) \log(\frac{f(x)}{g(x)}) dx$ .<sup>1</sup>

**Theorem 4** (Gibbs inequality). *The relative entropy is non-negative:  $D[p, q] \geq 0$ .*

*Proof.* Since  $-\log$  is convex, we get from the Jensen inequality  $D(p, q) = -\sum_k p_k \log(\frac{q_k}{p_k}) \geq -\log(\sum_k p_k \frac{q_k}{p_k}) = \log(\sum_k q_k) = \log(1) = 0$ .  $\square$

**12.9.** For  $p_k = 1/n$ , the uniform distribution, then entropy is  $\log(n)$ . It is an multi-variable calculus verification that  $0 \leq S(p) \leq \log(n)$ , where  $S(p) = 0$  in the deterministic case where only one  $p_k = 1$  and the others are zero. The uniform distribution has maximal entropy.

**12.10.** Remark. Here is an experimental observation for which I do not know the answer. Pick a large  $n$  and random  $p_k$  a random permutation  $\pi \in S_n$  and  $q_k = p_{\pi(k)}$ , then the Kullback-Leibler divergence  $D(p, q) = O(1)$ . For  $n \rightarrow \infty$  it seems to converge. For 30'000'000 we measure it to be around 3.29.

<sup>1</sup>Entropy and relative entropy is not defined for some perfectly nice bounded smooth distributions!