

Foundations of Probability Theory (MVE140 – MSA150)

Saturday 18th of January 2020 examination questions

You are allowed to use a dictionary (to and from English) and up to a maximum of one double-sided page of your own handwritten notes. This examination has five problems with a maximum of 20 credit points for a fully satisfactory solution, so the maximal total is 100 credit points. To pass the course, you need to score at least 40 points (50 if you are a PhD student). A member of staff is available at the examination site around 10:30am and the noon.

Examination Questions

1. A poker player receives 5 cards from a standard card deck containing 52 cards. What is the probability that he gets 4 cards of a kind, e.g., 4 aces or 4 kings, etc.? Do you think that the player is cheating if he shows twice such a situation in five consecutive games? (Take 0.001% as the cutoff point for your belief, so that you would not believe that the game is fair if you observe a too rare event, i.e. the probability of which is below 0.001%).

Solution. 13 variants to choose the kind and 48 variants for the other card. So $p = 13 \cdot 48 / \binom{52}{5} = 0.00024$. Probability of having twice such a situation in 5 games is binomial: $\binom{5}{2} p^2 (1-p)^{5-2} = 5.76 \cdot 10^{-7} < 10^{-5}$ so the player must be cheating.

2. n people at a supermarket, including Mr. X and Ms. Y, rush towards a newly opened cash till and get into the queue in the order of their arrival (i.e. randomly). What is the probability that between Mr. X and Ms. Y will be exactly $k = 0, \dots, n-2$ people standing in the queue?

Solution. There are $n!$ ways to place n people into the queue. There are $n-k-1$ positions in the queue separated by k other people and there are 2 ways to place X and Y in these places. Other $n-2$ people can be distributed over the remaining $n-2$ places in $(n-2)!$ ways. Thus the probability is $2(n-k-1)(n-2)!/n! = 2(1-k/(n-1))/n$.

3. An *entropy* of a discrete random variable ξ (or of its distribution) taking values x_i with probabilities p_i is $H(\xi) = -\sum_i p_i \log p_i$ (often \log_2 is considered). It is a measure of information received about the distribution when the random variable is sampled. Notice that it depends on the atom weights and not on their positions. Given two discrete random variables ξ, η with the joint distribution $\mathbf{P}\{\xi = x_i, \eta = y_j\} = p_{ij}$, their *joint entropy* is

$$H(\xi, \eta) = -\sum_{i,j} p_{ij} \log p_{ij}.$$

- (a) Compute $H(\beta)$ for Bernoulli $\beta \sim \text{Bern}(p)$ and find p for which $H(\beta)$ is minimal.
 (b) Show that if ξ and η are independent, then $H(\xi, \eta) = H(\xi) + H(\eta)$.
 (c) For discrete ξ, η , the *conditional entropy* of η with respect to the event $\{\xi = x_i\}$ is defined as

$$H_{x_i}(\eta) = -\sum_j \mathbf{P}\{\eta = y_j \mid \xi = x_i\} \log \mathbf{P}\{\eta = y_j \mid \xi = x_i\}$$

and the *mean conditional entropy* is

$$H_\xi(\eta) = \sum_i H_{x_i}(\eta) \mathbf{P}\{\xi = x_i\}.$$

Show that $H(\xi, \eta) = H(\xi) + H_\xi(\eta)$ and

- (d) $0 \leq H_\xi(\eta) \leq H(\eta)$.

Solution. (a) $H(\beta) = p \log p + (1-p) \log(1-p)$, differentiating w.r.t. p twice, the first derivative is 0 at $p = 1/2$, but the second is negative. Thus $H(\beta)$ is concave with maximum at $p = 1/2$ and minimum 0 at $p = 0$ and $p = 1$.

(b) Denoting $q_j = P\{\eta = y_j\}$, by independence $p_{ij} = p_i q_j$, and thus

$$\begin{aligned} H(\xi, \eta) &= - \sum_{i,j} p_i q_j (\log p_i + \log q_j) \\ &= - \sum_i p_i \log p_i \sum_j q_j - \sum_j q_j \log q_j \sum_i p_i = H(\xi) + H(\eta) \end{aligned}$$

(c) By the Joint probability formula,

$$H(\xi) + H_\xi(\eta) = - \sum_i p_i \log p_i - \sum_i p_i \sum_j \frac{p_{ij}}{p_i} (\log p_{ij} - \log p_i) = H(\xi, \eta)$$

(d) Non-negativity is obvious. Note that the function $h(x) = x \log x$, $x > 0$, is convex, therefore $\sum_i h(p_i) \geq h(\sum_i p_i)$ for any positive (p_i) such that $\sum_i p_i = 1$ (Jensen's inequality). Therefore,

$$\begin{aligned} H_\xi(\eta) &= - \sum_{i,j} p_i h(\mathbf{P}\{\eta = y_j \mid \xi = x_i\}) \\ &\leq - \sum_j h\left(\sum_i p_i \mathbf{P}\{\eta = y_j \mid \xi = x_i\}\right) \\ &= - \sum_j h(\mathbf{P}\{\eta = y_j\}) = H(\eta), \end{aligned}$$

where we have used the Full Probability formula for the argument of h .

4. $n \geq 2$ points U_1, \dots, U_n are thrown uniformly and independently on $[0, 1]$. Denote $\eta_1 = \min_{1 \leq k \leq n} U_k$ and $\eta_2 = \max_{1 \leq k \leq n} U_k$. Find:

- a) The c.d.f. of η_1 ;
- b) The density of η_2 ;
- c) The joint distribution of the pair (η_1, η_2) ;
- d) The conditional density of η_2 given $\eta_1 = x$, $x \in [0, 1]$.

Solution. a) $\eta_1 > x$, for $x \in [0, 1]$ when all the points u_k are inside $[x, 1]$, so by the independence, $F_{\eta_1}(x) = 1 - (1 - x)^n$. Obviously, it is 0 for $x < 0$ and 1 for $x > 1$.

b) Similarly, $F_{\eta_2}(x) = x^n$ since $\eta_2 \leq x$ iff all $u_k \leq x$. Thus $f_{\eta_2}(x) = nx^{n-1}$.

c) When $x > y$, $\mathbf{P}\{\eta_1 \leq x; \eta_2 \leq y\} = \mathbf{P}\{\eta_2 \leq y\} = F_{\eta_2}(y)$. For $0 \leq x \leq y \leq 1$, $\mathbf{P}\{x < \eta_1 \leq \eta_2 \leq y\} = (y - x)^n$, so that

$$F_{(\eta_1, \eta_2)}(x, y) = \mathbf{P}\{\eta_2 \leq y\} - \mathbf{P}\{x < \eta_1 \leq \eta_2 \leq y\} = y^n - (y - x)^n.$$

So the p.d.f. is

$$f_{(\eta_1, \eta_2)}(x, y) = \frac{\partial^2}{\partial x \partial y} F_{(\eta_1, \eta_2)}(x, y) = n(n-1)(y-x)^{n-2}$$

when $0 \leq x \leq y \leq 1$ and 0 otherwise.

d)

$$f_{\eta_2|\eta_1=x}(y) = \frac{f_{(\eta_1, \eta_2)}(x, y)}{f_{\eta_1}(x)} = \frac{(n-1)(y-x)^{n-2}}{(1-x)^{n-1}}$$

when $0 \leq x \leq y \leq 1$ and 0 otherwise.

5. Let $\mu_n = \max\{\xi_1, \dots, \xi_n\}$, where ξ_1, \dots, ξ_n are independent uniformly distributed on $[0, 1]$ random variables. Show that $n(1 - \mu_n)$ converges weakly to the Exponentially distributed random variable with parameter 1 (i.e. with the cdf $F(x) = 1 - e^{-x}$, $x \geq 0$).

Solution. For any $y \in [0, 1)$, $\mathbf{P}\{1 - \mu_n > y\} = \mathbf{P}\{\max\{\xi_1, \dots, \xi_n\} < 1 - y\} = \mathbf{P}\left(\bigcap_{k=1}^n \{\xi_k < 1 - y\}\right) = (1 - y)^n$. Therefore, for any $x \geq 0$ and any $n > x$ we will have $0 \leq x/n < 1$ and thus $\mathbf{P}\{n(1 - \mu_n) > x\} = (1 - x/n)^n \rightarrow e^{-x}$. The latter is $1 - F(x)$ for exponential distribution which is a continuous function. So the limit for all x means the weak convergence.