PHYSICS 210B : NONEQUILIBRIUM STATISTICAL PHYSICS
HW SOLUTIONS #1 : PROBABILITY

(1) A six-sided die is loaded in such a way that it is twice as likely to yield an even number than an odd number when thrown.

(a) Find the distribution $\{p_n\}$ consistent with maximum entropy.

(b) Assuming the maximum entropy distribution, what is the probability that three consecutive rolls of this die will total up to seven?

Solution:

(a) Our constraints are

$$\begin{aligned} p_1 + p_2 + p_3 + p_4 + p_5 + p_6 &= 1 \\ 2p_1 - p_2 + 2p_3 - p_4 + 2p_5 - p_6 &= 0 \ . \end{aligned}$$

We can combine these to yield

$$\begin{aligned} p_1 + p_3 + p_5 &= \tfrac{1}{3} \\ p_2 + p_4 + p_6 &= \tfrac{2}{3} \ . \end{aligned}$$

At this point it should be obvious that the solution is $p_{1,3,5} = \frac{1}{9}$ and $p_{2,4,6} = \frac{2}{9}$, since nothing further distinguishes among the even or the odd rolls. This is indeed what the maximum entropy construction gives. We write

$$S^*\big(\{p_n\}, \lambda_{\mathsf{o}}, \lambda_{\mathsf{e}}\big) = -\sum_{n=1}^{6} p_n \ln p_n - \lambda_{\mathsf{O}}\big(p_1 + p_3 + p_5 - \tfrac{1}{3}\big) - \lambda_{\mathsf{E}}\big(p_2 + p_4 + p_6 - \tfrac{2}{3}\big) \ .$$

Extremizing with respect to each of the six p_n, we have

$$\begin{aligned} p_1 = p_3 = p_5 &= e^{-(1+\lambda_{\mathsf{O}})} \\ p_2 = p_4 = p_6 &= e^{-(1+\lambda_{\mathsf{E}})} \ . \end{aligned}$$

Extremizing with respect to $\lambda_{\mathsf{O,E}}$ recovers the constraint equations. The solution is what we expected.

(b) There are 15 out of a possible $6^3 = 216$ distinct triples of die rolls which will total to seven:

$(1,1,5)$	$(2,1,4)$	$(3,1,3)$	$(4,1,2)$	$(5,1,1)$
$(1,2,4)$	$(2,2,3)$	$(3,2,2)$	$(4,2,1)$	
$(1,3,3)$	$(2,3,2)$	$(3,3,1)$		
$(1,4,2)$	$(2,4,1)$			
$(1,5,1)$				

Of these, six contain three odd rolls and nine contain one odd and two even rolls. Thus, the probability for three consecutive rolls summing to seven is

$$\pi = 6\,p_1^3 + 9\,p_1\,p_2^2 = \tfrac{14}{243} = 0.05761\,.$$

For a fair die the probability would be $\pi_{\text{fair}} = \frac{15}{216} = 0.06944$.

(2) Show that the Poisson distribution $P_\nu(n) = \frac{1}{n!}\,\nu^n\,e^{-\nu}$ for the discrete variable $n \in \mathbb{Z}_{\geq 0}$ tends to a Gaussian in the limit $\nu \to \infty$.

Solution:

For large fixed ν, $P_\nu(n)$ is maximized for $n \sim \nu$. We can see this from Stirling's asymptotic expression,

$$\ln n! = n\ln n - n + \tfrac{1}{2}\ln n + \tfrac{1}{2}\ln 2\pi + \mathcal{O}(1/n)\ ,$$

which yields

$$\ln P_\nu(n) = n\ln\nu - n\ln n - \nu + n - \tfrac{1}{2}\ln n - \tfrac{1}{2}\ln 2\pi$$

up to terms of order $1/n$, which we will drop. Varying with respect to n, which we can treat as continuous when it is very large, we find $n = \nu - \frac{1}{2} + \mathcal{O}(1/\nu)$. We therefore write $n = \nu + \frac{1}{2} + \varepsilon$ and expand in powers of ε. It is easier to expand in powers of $\tilde\varepsilon \equiv \varepsilon + \frac{1}{2}$, and since n is an integer anyway, this is really just as good. We have

$$\ln P_\nu(\nu + \tilde\varepsilon) = -(\nu + \tilde\varepsilon)\ln\!\left(1 + \frac{\tilde\varepsilon}{\nu}\right) + \tilde\varepsilon - \tfrac{1}{2}\ln(\nu + \tilde\varepsilon) - \tfrac{1}{2}\ln 2\pi\ .$$

Now expand, recalling $\ln(1+z) = z - \frac{1}{2}z^2 + \ldots$, and find

$$\ln P_\nu(\nu + \tilde\varepsilon) = -\frac{\tilde\varepsilon(1+\tilde\varepsilon)}{2\nu} - \ln\sqrt{2\pi\nu} + \frac{\tilde\varepsilon^2}{4\nu^2} + \ldots$$

Since $\nu \to \infty$, the last term before the ellipses is negligible compared with the others, assuming $\tilde\varepsilon = \mathcal{O}(\nu^0)$. Thus,

$$P_\nu(n) \sim (2\pi\nu)^{-1/2}\exp\left\{-\frac{\left(n - \nu + \frac{1}{2}\right)^2}{2\nu}\right\}\ ,$$

which is a Gaussian.

(3) The probability density for a random variable x is given by the Lorentzian,

$$P(x) = \frac{\gamma}{\pi}\cdot\frac{1}{x^2 + \gamma^2}\ .$$

Consider the sum $X = \sum_{i=1}^N x_i$, where each x_i is independently distributed according to $P(x_i)$.

(a) Find the distribution $P_N(X)$. Does it satisfy the central limit theorem? Why or why not?

(b) Find the probability $\varPi_N(Y)$ that $|X_N| < Y$, where $Y > 0$ is arbitrary.

Solution :

(a) As discussed in the Lecture Notes, the distribution of a sum of identically distributed random variables, $X = \sum_{i=1}^{N} x_i$, is given by

$$P_N(X) = \int\limits_{-\infty}^{\infty} \frac{dk}{2\pi} \left[\hat{P}(k)\right]^N e^{ikX} ,$$

where $\hat{P}(k)$ is the Fourier transform of the probability distribution $P(x_i)$ for each of the x_i. The Fourier transform of a Lorentzian is an exponential:

$$\int\limits_{-\infty}^{\infty} dx\, P(x)\, e^{-ikx} = e^{-\gamma|k|} .$$

Thus,

$$\begin{aligned} P_N(X) &= \int\limits_{-\infty}^{\infty} \frac{dk}{2\pi}\, e^{-N\gamma|k|}\, e^{ikX} \\ &= \frac{N\gamma}{\pi} \cdot \frac{1}{X^2 + N^2\gamma^2} . \end{aligned}$$

This is not a Gaussian. The central limit theorem does not apply because the Lorentzian distribution has no finite second moment.

(b) The probability for X to lie in the interval $X \in [-Y, Y]$, where $Y > 0$, is

$$\varPi_N(Y) = \int\limits_{-Y}^{Y} dX\, P_N(X) = \frac{2}{\pi}\, \tan^{-1}\!\left(\frac{Y}{N\gamma}\right) .$$

The integral is easily performed with the substitution $X = N\gamma \tan\theta$. Note that $\varPi_N(0) = 0$ and $\varPi_N(\infty) = 1$.

(4) Frequentist and Bayesian statistics can sometimes lead to different conclusions. You have a coin of unknown origin. You assume that flipping the coin is a Bernoulli process, *i.e.* the flips are independent and each flip has a probability p to end up heads and probability $1 - p$ to end up tails.

(a) You perform 14 flips of the coin and you observe the sequence {HHTHTHHHTTHHHH}. As a frequentist, what is your estimate of p?

(b) What is your frequentist estimate for the probability that the next two flips will each end up heads? Would you bet on this event?

(c) Now suppose you are a Bayesian. You view p as having its own distribution. The likelihood $f(\mathrm{data}|p)$ is still given by the Bernoulli distribution with the parameter p. For the prior $\pi(p)$, assume a Beta distribution,

$$\pi(p|\alpha,\beta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\,\Gamma(\beta)}\,p^{\alpha-1}\,(1-p)^{\beta-1} \quad .$$

where α and β are hyperparameters. Compute the posterior distribution $\pi(p|\mathrm{data},\alpha,\beta)$.

(d) What is the posterior predictive probability $f(\mathsf{HH}|\mathrm{data},\alpha,\beta)$?

(e) Since *a priori* we don't know anything about the coin, it seems sensible to choose $\alpha=\beta=1$ initially, corresponding to a flat prior for p. What is the numerical value of the probability to get two heads in a row? Would you bet on it?

Solution:

(a) A frequentist would conclude $p=\frac{5}{7}$ since the trial produced ten heads and four tails.

(b) The frequentist reasons that the probability of two consecutive heads is $p^2=\frac{25}{49}$. This is slightly larger than $\frac{1}{2}$, so the frequentist should bet! (Frequently, in fact.)

(c) Are you reading the lecture notes? You should, because this problem is solved there in §1.5.2. We have

$$\pi(p|\mathrm{data},\alpha,\beta) = \frac{p^{9+\alpha}\,(1-p)^{3+\beta}}{\mathsf{B}(10+\alpha,4+\beta)}\,,$$

where the Beta function is $\mathsf{B}(\alpha,\beta)=\Gamma(\alpha)\,\Gamma(\beta)/\Gamma(\alpha+\beta)$.

(d) The posterior predictive is

$$p(\mathrm{data}'|\mathrm{data}) = \frac{\mathsf{B}(10+Y+\alpha,4+M-Y+\beta)}{\mathsf{B}(10+\alpha,4+\beta)}\,,$$

where Y is the total number of heads found among M tosses. We are asked to consider $M=2$, $Y=2$, so

$$f(\mathsf{HH}|\mathrm{data},\alpha,\beta) = \frac{\mathsf{B}(12+\alpha,4+\beta)}{\mathsf{B}(10+\alpha,4+\beta)}\,.$$

(e) With $\alpha=\beta=1$, we have

$$f(\mathsf{HH}|\mathrm{data},\alpha,\beta)\Big|_{\alpha=\beta=1} = \frac{\mathsf{B}(13,5)}{\mathsf{B}(11,5)} = \frac{11\cdot 12}{16\cdot 17} = \frac{33}{68} = 0.4852941.$$

This is slightly less than $\frac{1}{2}$. Don't bet!

It is instructive to note that the Bayesian posterior prediction for a single head, assuming $\alpha=\beta=1$, is

$$f(\mathsf{H}|\mathrm{data},\alpha,\beta)\Big|_{\alpha=\beta=1} = \frac{\mathsf{B}(11+\alpha,4+\beta)}{\mathsf{B}(10+\alpha,4+\beta)} = \frac{\mathsf{B}(12,5)}{\mathsf{B}(11,5)} = \frac{11}{16}\,.$$

The square of this number is $\frac{121}{256} = 0.4726565$, which is less than the posterior prediction for two consecutive heads, *even though our likelihood function is the Bernoulli distribution, which assumes the tosses are statistically independent*. The eager student should contemplate why this is the case.

(5) Consider the family of distributions

$$f(\boldsymbol{k}|\lambda) = \prod_{j=1}^{N} \frac{\lambda^{k_j}\, e^{-\lambda}}{k_j!} ,$$

corresponding to a set of N independent discrete events characterized by a Poisson process with Poisson parameter λ. Show that

$$\pi(\lambda|\alpha,\beta) = \frac{\alpha^\beta}{\Gamma(\beta)}\, \lambda^{\beta-1}\, e^{-\alpha\lambda} ,$$

is a family of priors, each normalized on $\lambda \in [0,\infty)$, which is conjugate to the likelihood distributions $f(\boldsymbol{k}|\lambda)$.

Solution:

Let $K = k_1 + \cdots + k_N$. Then

$$\begin{aligned}
\pi(\lambda|\boldsymbol{k},\alpha,\beta) &= \frac{f(\boldsymbol{k}|\lambda)\,\pi(\lambda|\alpha,\beta)}{\int\limits_0^1 d\lambda'\, f(\boldsymbol{k}|\lambda')\,\pi(\lambda'|\alpha,\beta)} \\
&= \frac{\lambda^{K+\beta-1}\, e^{-(N+\alpha)\lambda}}{\int\limits_0^1 d\lambda'\, {\lambda'}^{K+\beta-1}\, e^{-(N+\alpha)\lambda'}} = \frac{(N+\alpha)^{K+\beta}}{\Gamma(K+\beta)}\, \lambda^{K+\beta-1}\, e^{-(N+\alpha)\lambda} .
\end{aligned}$$

Thus the change in hyperparameters is given by

$$\begin{aligned}
\alpha \to \alpha' &= N + \alpha \\
\beta \to \beta' &= K + \beta .
\end{aligned}$$

(6) Professor Jones begins his academic career full of hope that his postdoctoral work, on relativistic corrections to the band structure of crystalline astatine under high pressure, will eventually be recognized with a Nobel Prize in Physics. Being of Bayesian convictions, Jones initially assumes he will win the prize with probability θ, where θ is uniformly distributed on $[0,1]$ to reflect Jones' ignorance.

(a) After N years of failing to win the prize, compute Jones's chances to win in year $N+1$ by performing a Bayesian update on his prior distribution.

(b) Jones' graduate student points out that Jones' prior is not parameterization-independent. He suggests Jones redo his calculations, assuming initially the Jeffreys prior for the Bernoulli process. What then are Jones' chances after his N year drought?

(c) Professor Smith, of the Economics Department, joined the faculty the same year as Jones. His graduate research, which concluded that poor people have less purchasing power than rich people, was recognized with a Nobel Prize in his fifth year. Like Jones, Smith is a Bayesian, whose initial prior distribution was taken to be uniform. What is the probability he will win a second Nobel Prize in year 11? If instead Smith were a frequentist, how would he assess his chances in year 11?

Solution:

(a) For the Beta distribution $\pi(\theta) = \theta^{\alpha-1}(1-\theta)^{\beta-1}/B(\alpha,\beta)$, one has

$$\langle\theta\rangle = \frac{\alpha}{\alpha+\beta} \quad . \tag{1}$$

Assuming $\alpha_0 = \beta_0 = 1$, under the Bayesian update rules, $\alpha_N = \alpha + P$ and $\beta_N = \beta + N - P$, where P is the number of successes in N years. Alas, for Jones $P = 0$, so $\alpha_N = 1$ and $\beta_N = N + 1$, meaning $f(\textsf{prize}|\textsf{reality}) = 1/(N+2)$.

(b) For the Jeffries prior, take $\alpha_0 = \beta_0 = \frac{1}{2}$, in which case $f(\textsf{prize}|\textsf{reality}) = 1/(2N+2)$.

(c) For Smith, we take $P = 1$ and $N = 10$, hence $f(\textsf{prize}|\textsf{reality}) = 2/(N+2) = \frac{1}{6}$. If Smith were a frequentist, he would estimate his chances at $p = \frac{1}{10}$.