Review Session Notes

Feng Zhao

June 22, 2021

1 Probability Theory Exercise 1

1. For multiple random variables, pairwise independence does not imply joint independence: Let X, Y be independent Bernoulli random variables with parameter $\frac{1}{2}$, and $Z = X \oplus Y$. $P(X = 1, Y = 1, Z = 1) = 0$ (See 1-1)

2. Using formulas of conditional probability $P(A|B)$, Bayes formula, addition formula for disjoint event and multiplication formula for independent events. (See 1-2,1-3,2-1)

3. omit

4. Given a finite sample space Ω and a subset $C \subset 2^\Omega$, compute the size of the $\sigma-$ field generated from C, written as $\sigma(C)$. The general formula for $C = \{A_1, \dots, A_k\}$ is 2^k if $\cup_{i=1}^k = \Omega$ and $A_i \cap A_j = \emptyset$. (See 1-4 [1])

   (a) $\sigma(C) = 16$. Notice that $\Omega = \{HH, HT, TT, TH\}$ which is equivalent with $\{1, 2, 3, 4\}$. Further $A = \{1, 2\}$, $B = \{1, 4\}$, $\mathcal{C} = \{A, B\}$. Then $|\sigma(\mathcal{C})| = |\sigma(\{\{1\}, \{2\}, \{3\}, \{4\}\})| = 2^4 = 16$.

   (b) $\Omega = \{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT\}$ which is equivalent to $\{1, 2, 3, 4, 5, 6, 7, 8\}$, $C = \{\{1, 2, 7\}, \{1, 6, 7\}, \{1, 2, 6\}\}$ $|\mathcal{C}| = |\sigma(\{\{1\}, \{2\}, \{6\}, \{7\}, \{3, 4, 5, 8\}\})| = 2^5 = 32$.

5. BC-Lemma proof technique, definition of $\limsup_{n\to\infty} A_n := \cap_{k=1}^\infty \cup_{n\geq k} A_n$. $\mathbb{P}(\cap_{i=1}^\infty \cup_{n=i}^\infty A_n) = 1$. Proof: From the condition $\mathbb{P}(\cup_{n=1}^\infty A_n) = 1$ follows $\mathbb{P}(\cap_{n=1}^\infty A_n^c) = 0$. Since $\{A_n\}_{n=1}^\infty$ is an independent sequence, we have $\prod_{n=1}^\infty P(A_n^c) = 0$. Since $\mathbb{P}(A_n) < 1$, $\mathbb{P}(A_n^c) > 0$. Then for any i, $\prod_{n=i}^\infty P(A_n^c) = 0 \Rightarrow \mathbb{P}(\cap_{n=i}^\infty A_n^c) = 0$. The probability $\mathbb{P}(\cup_{i=1}^\infty \cap_{n=i}^\infty A_n^c) \leq \sum_{i=1}^\infty \mathbb{P}(\cap_{n=i}^\infty A_n^c) = 0$. Taking the complement we get $\mathbb{P}(\cap_{i=1}^\infty \cup_{n=i}^\infty A_n) = 1$ at last.

6. Definition of random variables, mapping from $\Omega \to \mathbb{R}$ which satisfies additional condition: If $X_1, X_2, \dots$ are random variables, and X_n converges pointwise to X, show that $X_1 + X_2, X$ are also random variables.

[1] exercise 1, 4th problem

(a) Since $\{w|X_1(w)+X_2(w)<c\}=\bigcup_{r\in Q}\{w|X_1(w)>r\}\cap\{w|X_2(w)<c-r\}$ is F-measurable, X_1+X_2 is a random variable.

(b) $\{w|\lim_{n\to\infty}X_n(w)\le c\}=\bigcap_{n=1}^{\infty}\bigcup_{i=1}^{\infty}\bigcap_{j=i}^{\infty}\{w|X_n(w)<c+\frac{1}{n}\}$ is F-measurable $\Rightarrow \lim_{n\to\infty}X_n$ is a random variable.

2 Probability Theory Exercise 2

1. omit

2. Maximum entropy distribution.

   (a) Using the log sum inequality we have $\sum_{i=1}^{n}p_i\log\frac{p_i}{Cr^u}\ge 0$, which is equivalent to $\sum_{i=1}^{n}p_i\log p_i\ge\log C+\mu\log r=\sum_{i=1}^{n}Cr^{x_i}\log(Cr^{x_i})$, since C,r satisfy the two constraints $\sum_{i=1}^{n}Cr^{x_i}=1$ and$\sum_{i=1}^{n}Cr^{x_i}x_i=\mu$.
   Another method: using Lagrange multiplier to minimize $\sum_{k=1}^{n}p_k\log p_k-\lambda_1(\sum_{k=1}^{n}p_k-1)-\lambda_2(\sum_{k=1}^{n}p_kx_k-\mu)$. Taking the partial derivative on p_k, we get $p_k=\exp(\lambda_1+\lambda_2x_k+1)$. Let $C=\exp(\lambda_1+1)$ and $r=e^{\lambda_2}$. we get the required distribution $p_k=Cr^{x_k}$.

   (b) $p_i=Cr^i$, which has the same PMF as the geometric distribution. Let p be the parameter of the geometric distribution. Then we have $\mu=\frac{1}{p}$, $C=\frac{p}{1-p}=\frac{1}{\mu-1}$ and $r=1-p=1-\frac{1}{\mu}$.

3. Series sum, using $S_n=\sum_{i=1}^{n}\frac{1}{n}\sim\log n+\gamma$.

   (a) Let the partial sum $T_n=\sum_{i=1}^{n}\frac{(-1)^{i+1}}{i}$. Also we define $S_n=\sum_{i=1}^{n}\frac{1}{n}$. Then we can show that $T_{2n}=S_{2n}-S_n$. Using the asymptotic form for harmonic series, we have $S_n\sim\log n+\gamma$ where γ is the Euler constant. Therefore, $T_{2n}\sim\log(2n)-\log n=\log 2$. That is $\sum_{i=1}^{\infty}\frac{(-1)^{n+1}}{n}=\log 2$.

   (b) Using similar method as above, let $F_{3n}=\sum_{i=1}^{n}(\frac{1}{2i-1}-\frac{1}{4i}-\frac{1}{4i-2})$. We will show that $\lim_{n\to\infty}F_{3n}=\frac{1}{2}\log 2$. It is obvious that $F_{3n}=\frac{1}{2}\sum_{i=1}^{n}(\frac{1}{2i-1}-\frac{1}{2i})=\frac{1}{2}T_n$. Using conclusion from (a) we reach the conclusion.

   (c) $\sum_{i=1}^{\infty}\frac{1}{n}=\infty$.

4. higher order moments for geometric and Poisson distribution. For geometric distribution, we have $\mathbb{E}[(X-1)^k|X>1]=\mathbb{E}[X^k]$; For Poisson distribution, we have $\mathbb{E}[X(X-1)\dots(X-k)]=\lambda^{k+1}$. If $X\sim$ Geometric(p). we already know that $\mathbb{E}[X]=\frac{1}{p},\mathbb{E}[X^2]=\frac{2-p}{p^2}$. For third order moment,

$$\begin{aligned}\mathbb{E}[X^3]&=P(X=1)\,\mathbb{E}[X^3|X=1]+P(X>1)\,\mathbb{E}[X^3|X>1]\\&=p+(1-p)(\mathbb{E}[(X-1)^3|X>1]+3\,\mathbb{E}[(X-1)^2|X>1]+3\,\mathbb{E}[(X-1)|X>1]+1)\\&=p+(1-p)(\mathbb{E}[X^3]+\frac{3(2-p)}{p^2}+\frac{3}{p}+1)\end{aligned}$$

Solving the equation we can get $\mathbb{E}[X^3] = \frac{p^2+6-6p}{p^3}$. For Poisson distribution, we already known that $\mathbb{E}[X] = \lambda, \mathbb{E}[X^2] = \lambda^2 + \lambda$. Using the formula $\mathbb{E}[X(X-1)(X-2)] = \lambda^3$ we can solve out $\mathbb{E}[X^3] = \lambda^3 + 3\lambda^2 + \lambda$.

5. Using Stirling's formula, $n! \sim \sqrt{2\pi n}(\frac{n}{e})^n$.

   $\beta_0 = \frac{1}{2}$. Let A represents the event that no collision happens. Then $P(A) = \frac{(n-1)\ldots(n-k+1)}{n^{k-1}}$. We analyze the asymptotic behaviour of $P(A)$ by Stirling's formula as follows: $P(A) = \frac{n!}{n^k(n-k)!} \sim \frac{(n/e)^n}{n^k((n-k)/e)^{n-k}}$. We can write $\frac{(n/e)^n}{n^k((n-k)/e)^{n-k}} = \frac{1}{\exp(k+(n-k)\log(1-\frac{k}{n}))} = \frac{1}{\exp((n+k)k^2/(2n^2)+O(k^3/n^2))}$. If $\beta < \frac{1}{2}$, $k^2 < n$ and $P(A) \to 1$. Otherwise $P(A) \to 0$.

   Another method. We can write $P(A)$ as $\prod_{i=1}^{k=1} \frac{n-i}{n}$, $P(A)$ converges to 1 or 0 is equivalent to the series $\sum_{i=1}^{k-1} \log(1-\frac{i}{n})$ converges to 0 or diverges to $-\infty$. To analysis the series, we use the inequality $x - \frac{x^2}{2} < \log(1+x) < x$ for $x < 0$. Then

$$-\left(\sum_{i=1}^{k-1} \frac{i}{n} + \sum_{i=1}^{k-1} \frac{i^2}{2n^2}\right) < \sum_{i=1}^{k-1} \log(1-\frac{i}{n}) < -\sum_{i=1}^{k-1} \frac{i}{n}$$
$$\Rightarrow -\left(\frac{k(k-1)}{2n} + O(\frac{k^3}{n^2})\right) < \sum_{i=1}^{k-1} \log(1-\frac{i}{n}) < -\frac{k(k-1)}{2n}$$

   Since $k = \lceil n^\beta \rceil$, $\frac{k(k-1)}{2n} \sim \frac{1}{2}n^{2\beta-1}$. When $\beta < \frac{1}{2}$, both the upper bound and lower bound of $\sum_{i=1}^{k-1} \log(1-\frac{i}{n})$ converge to 0, which corresponds to $P(A) = 1$. On the other hand, when $\beta > \frac{1}{2}$, both sides diverge to $-\infty$.

6. Cauchy criterion.

   (a) The recursive formula is

$$f(m) = \frac{1}{6}\sum_{i=1}^{6} f(m-i) \tag{1}$$

   so that we get $f(2020) \approx 0.286$ by computer program.

   (b) The complement of the event $\exists n, \text{s.t.} S_n = m$ is $\exists n, \text{s.t.} S_n < m \wedge S_{n+1} \geq m$, which can be further decomposed into 5 events: $S_n = m - j + 1 \wedge X_{n+1} \geq j$ for $j = 2, 3, 4, 5, 6$. Therefore, we have the following equation: $1 - f(m) = \sum_{j=1}^{5} \frac{6-i}{6} f(m-j)$. Taking the limit $m \to \infty$ on both sides we have $\lim_{m\to\infty} f(m) = \frac{2}{7}$.

   (c) Let $f(0) = 1$. Then $f(6)$ satisfies the recursive formula (1). Since $0 \leq f(m) \leq 1$, then $\max_m |f(m-j) - f(m)| \leq 1$ for $m \geq 0$ and $j = 1, 2, \ldots, 5$. Using the recursive formula, first we have $|f(m) - f(m-j)| = \frac{1}{6}|\sum_{i=1}^{6}[f(m-i) - f(m-j)]| \leq \frac{5}{6}$ for $m \geq 6$ and $j = 1, 2, \ldots, 5$. Then we have $|f(m) - f(m-j)| = \frac{1}{6}|\sum_{i=1}^{6}[f(m -$

$i) - f(m-j)]| \le (\frac{5}{6})^2$ for $m \ge 12$ and $j = 1, 2, \ldots, 5$. Recursively we have $[f(m) - f(m-j)]| \le (\frac{5}{6})^{\lfloor m/6 \rfloor}$ for $j = 1, 2, \ldots, 5$. Then $|f(m+n) - f(m)| \le \sum_{i=1}^{n} |f(m+i) - f(m+i-1)| \le \sum_{i=1}^{n} (\frac{5}{6})^{\lfloor (m+i)/6 \rfloor} \le 6\frac{1}{1-(5/6)}(\frac{5}{6})^{\lfloor (m+1)/6 \rfloor} = 36(\frac{5}{6})^{\lfloor (m+1)/6 \rfloor}$. By Cauchy's convergence test, $f(m)$ converges.

Another proof, analysis of complex root (linear difference equation). The characteristic equation of (1) is $6z^6 = 1+z+z^2+z^3+z^4+z^5$. We will show that this equation has only one root whose norm is 1, and the norm of all other roots is less than 1. Then $f(n)$ will converge to a constant. For $|z| > 1$, we have $|\frac{1}{z}| < 1$, by triangular inequality $|\frac{1}{z} + \cdots + \frac{1}{z^6}| < |\frac{1}{z}| + \cdots + |\frac{1}{z^6}| < 6$. Therefore, $\frac{1}{z} + \cdots + \frac{1}{z^6} \ne 6$ and $|z| > 1$ is not a root of $6z^6 = 1+z+z^2+z^3+z^4+z^5$. On the other hand, multiplying the equation on both sides by $z-1$ we have $6z^7 - 7z^6 + 1 = 0$. Assume $z = e^{i\theta}$ for some $0 \le \theta < 2\pi$ is a root of this transformed equation, we will show that $\theta = 0$. Indeed, we can write the equation system of the real part and the imaginary part as

$$
\begin{aligned}
6\cos 7\theta - 7\cos 6\theta + 1 &= 0 \\
6\sin 7\theta &= 7\sin 6\theta
\end{aligned}
$$

Using the identity $\sin^2 t + \cos^2 t = 1$ to cancel 7θ, we have $\cos 6\theta = 1$ and $\sin 7\theta = 0$, which only allows $\theta = 0$. Finally, we verify that $z = 1$ is not duplicate roots of $6z^6 = 1+z+z^2+z^3+z^4+z^5$ by verifying that $z = 1$ is not a root of $36z^5 = 1+2z+3z^2+4z^3+5z^4$ (derivative).

3 Probability Theory Exercise 3

1. Without the positive constraint, MCT does not hold. Let $(\Omega, \mathcal{F}, P) = ([0,1], \mathcal{B}, \lambda)$ where $\mathcal{B}$ is the Borel σ-algebra and λ is the Lebesgue measure. Let $X_n(w) = \frac{-1}{nw}$ for $w \in (0,1]$. Then $\mathbb{E}[X_n] = -\infty$ while $X_n \to X = 0$, whose expectation is zero. (See 3-1)

2. (a) $c = \frac{1}{2\pi}$.

   (b) When k is odd, $E[X^k] = 0$. When $k = 2m$ for $m = 2, 4, \ldots$, using the Beta function $B(x,y) = \int_0^1 t^{x-1}(1-t)^{y-1}dt$, we have

$$
\begin{aligned}
E[X^k] = c\int_{-2}^{2} x^{2m}\sqrt{4-x^2}dx \qquad &= \frac{1}{\pi}\int_0^1 x^{2m}\sqrt{4-x^2}dx \\
= 2^{2m+1}\int_0^1 t^{m-\frac{1}{2}}(1-t)^{1/2}dt \qquad &= \frac{2^{2m+1}}{\pi}B(m+\frac{1}{2}, \frac{3}{2}) \\
= \frac{2^{m+1}(2m-1)!!}{(m+1)!} \qquad &= \frac{2}{m+1}\frac{(2m)!}{(m!)^2}
\end{aligned}
$$

3. Make induction on n. Lemma 5 of [1]. For $n = 1$, $P(X \geq k) \geq P(Y \geq k)$ obviously holds. If the conclusion holds for $n = m$, then for $n = m + 1$,

$$\begin{aligned} P(\sum_{i=1}^{m+1} X_i \geq k) &= P(\sum_{i=1}^{m} X_i \geq k, X_{m+1} = 0) + P(\sum_{i=1}^{m} X_i \geq k-1, X_{m+1} = 1) \\ &= P(\sum_{i=1}^{m} X_i \geq k) + P(\sum_{i=1}^{m} X_i = k-1) P(X_{m+1} = 1) \end{aligned}$$

Similarly,

$$P(\sum_{i=1}^{m+1} Y_i \geq k) = P(\sum_{i=1}^{m} Y_i \geq k) + P(\sum_{i=1}^{m} Y_i = k-1) P(Y_{m+1} = 1)$$

If $P(\sum_{i=1}^{m} X_i = k-1) \geq P(\sum_{i=1}^{m} Y_i = k-1)$, then using induction for $n = m$ we have $P(\sum_{i=1}^{m+1} X_i \geq k) \geq P(\sum_{i=1}^{m+1} Y_i \geq k)$. Otherwise,

$$\begin{aligned} P(\sum_{i=1}^{m+1} X_i \geq k) - P(\sum_{i=1}^{m+1} Y_i \geq k) &= P(\sum_{i=1}^{m} X_i \geq k) - P(\sum_{i=1}^{m} Y_i \geq k) \\ &\quad + P(X_{m+1} = 1)(P(\sum_{i=1}^{m} X_i = k-1) - P(\sum_{i=1}^{m} Y_i = k-1)) \\ &\geq P(\sum_{i=1}^{m} X_i \geq k) - P(\sum_{i=1}^{m} Y_i \geq k) + (P(\sum_{i=1}^{m} X_i = k-1) - P(\sum_{i=1}^{m} Y_i = k-1)) \\ &= P(\sum_{i=1}^{m} X_i \geq k-1) - P(\sum_{i=1}^{m} Y_i \geq k-1) \underset{(a)}{\geq} 0 \end{aligned}$$

where (a) also follows from the deduction of $n = m$.

4. Multivariate change of variables

5. Calculus

6. Gaussian Orthogonal Ensemble. The joint probability distribution is

$$p(y_1, y_2) = c|y_1 - y_2| \exp(-\frac{y_1^2 + y_2^2}{4}) \tag{2}$$

Let $A = \begin{bmatrix} X_1 & X_3 \\ X_3 & X_2 \end{bmatrix}$, the characteristic equation of A is $\det(A - \lambda I) = (X_1 - \lambda)(X_2 - \lambda) - X_3^2 = 0$.

Y_1 and Y_2 are two roots of this eqution, we get $\begin{cases} Y_1 + Y_2 = X_1 + X_2 \\ Y_1 Y_2 = X_1 X_2 - X_3^2 \end{cases}$.

Also, $\begin{cases} Y_1 + Y_2 = X_1 + X_2 \\ (Y_1 - Y_2)^2 = 4((\frac{X_1 - X_2}{2})^2 + X_3^2) \end{cases}$. We have $X_1 \sim N(0, 2), X_2 \sim$

$N(0,2), X_3 \sim N(0,1)$.
So, $(X_1+X_2) \sim N(0,4)$, $\frac{X_1-X_2}{2} \sim N(0,1)$.
$(\frac{X_1-X_2}{2})^2 \sim \chi^2(1), X_3^2 \sim \chi^2(1)$, $((\frac{X_1-X_2}{2})^2 + X_3^2) \sim \chi^2(2) = \text{Exp}(\frac{1}{2})$.

We define $\begin{cases} Z_1 = Y_1 + Y_2 = X_1 + X_2 \\ 4Z_2 = (Y_1 - Y_2)^2 = (X_1 - X_2)^2 + X_3^2 \end{cases}$. Then, $Z_1 \sim N(0,4)$ and $Z_2 \sim \text{Exp}(\frac{1}{2})$. Since $\mathbb{E}[(X_1-X_2)(X_1+X_2)] = 0$, both X_1-X_2, X_1+X_2 follow Gaussian distribution, $X_1 - X_2 \perp\!\!\!\perp X_1 + X_2 \Rightarrow Z_1 \perp\!\!\!\perp Z_2$.
Let $Z_3 = Y_1 - Y_2$. Then for $z < 0$, $P(Z_3 < z) = \frac{1}{2}P(Z_2 > z^2/4) = \int_{z^2/4}^{+\infty} \frac{1}{4}e^{-u/2}du$. Taking the derivative on both sides, we have $f_3(z) = -\frac{z}{8}\exp(-\frac{z^2}{8})$ where f_3 is the pdf of Z_3 Similarly, for $z > 0$ we have $f_3(z) = \frac{z}{8}\exp(-\frac{z^2}{8})$. Therefore, the pdf of Z_3 is $f_3(z) = \frac{|z|}{8}\exp(-\frac{z^2}{8})$. $Z_1 \perp\!\!\!\perp Z_2 \Rightarrow Z_1 \perp\!\!\!\perp Z_3$. Now we consider the transformation from (Z_1, Z_3) to (Y_1, Y_2) The determinant of the Jacobian matrix is $\mid J \mid = \left| \frac{\partial(z_1,z_3)}{\partial(y_1,y_2)} \right| = 2$. Finally, we get

$$\begin{aligned} f_{Y_1,Y_2}(y_1,y_2) &= f_{Z_1,Z_3}(y_1+y_2, y_1-y_2) \mid J \mid \\ &= \frac{1}{2\sqrt{2\pi}} e^{-\frac{(y_1+y_2)^2}{8}} \frac{1}{8}|y_1-y_2| e^{-\frac{(y_1-y_2)^2}{8}} 2 \\ &= \frac{1}{8\sqrt{2\pi}} e^{\frac{-y_1^2-y_2^2}{4}} \mid y_1 - y_2 \mid \end{aligned} \tag{3}$$

Another method, the eigenvalue decomposition of $\begin{pmatrix} X_1 & X_3 \\ X_3 & X_2 \end{pmatrix}$ is given by:

$$\begin{pmatrix} X_1 & X_3 \\ X_3 & X_2 \end{pmatrix} = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} Y_1 & 0 \\ 0 & Y_2 \end{pmatrix} \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix}$$

After expansion, we have

$$\begin{aligned} X_1 &= Y_1 \cos^2\theta + Y_2 \sin^2\theta \\ X_2 &= Y_1 \sin^2\theta + Y_2 \cos^2\theta \\ X_3 &= (Y_1 - Y_2)\sin\theta\cos\theta \end{aligned}$$

The absolute value of the determinant of the transformation matrix is given by

$$|\frac{\partial(x_1,x_2,x_3)}{\partial(y_1,y_2,\theta)}| = f(\theta)|y_1 - y_2|$$

Therefore, the distribution of (Y_1, Y_2, θ) is given by

$$\begin{aligned} p(y_1,y_2,\theta) =& c' f(\theta)|y_1-y_2| \exp(-\frac{(y_1\cos^2\theta + y_2\sin^2\theta)^2 + (y_1\sin^2\theta + y_2\cos^2\theta)^2}{4}) \\ &\cdot \exp(-\frac{\sin^2\theta\cos^2\theta(y_1-y_2)^2}{2}) \\ =& c' f(\theta)|y_1-y_2| \exp(-\frac{y_1^2+y_2^2}{4}) \end{aligned}$$

where c' is a constant. After integration of θ over $[0, 2\pi]$, we can get the marginal distribution for (y_1, y_2) in (2).

4 Probability Theory Exercise 4

1. (a) If $s = 0$, $M_X(s) = 1$. Below we show that $M_X(s) = +\infty$. For any $s > 0$, we can find M satisfying $M > t$ and $e^{sx/2} > x^2$ for $x > M$. Then $M_X(s) = \int_{-\infty}^{+\infty} \frac{1}{\pi} \frac{t}{t^2+x^2} e^{sx} dx \geq \frac{t}{2\pi} \int_M^\infty \frac{e^{sx}}{x^2} dx \geq \frac{t}{2\pi} \int_M^\infty \frac{e^{sx/2}}{x^2} dx = +\infty$. Similar proof can be made for $s < 0$.

   (b) My answer is yes. The symmetric log-normal distribution satisfies the two conditions. Let $X = e^{-Z} + e^{Z}$ where Z is standard normal distribution. Then $E[X^n] = e^{-n^2/2} + e^{n^2/2}$, which is finite. On the other hand, $M_X(s) = +\infty$ for any $s \neq 0$.

2. (a) From $\log x \leq x - 1$, we have $e^{tx-k} \geq \frac{t^k x^k}{k^k}$. Taking the expectation on both sides, $M_X(t)e^{-k} \geq \frac{t^k}{k^k} \mathbb{E}[X^k]$. Let $t = a > 0$. We then have $\mathbb{E}[X^k] \leq (\frac{k}{ae})^k M_X(a) < +\infty$.

   (b) We use the same inequality $e^{tx-k} \geq \frac{t^k x^k}{k^k}$ as (a) but we let $t = a - s > 0$, we then have $e^{ax-k} \geq e^{sx} x^k (\frac{a-s}{k})^k$. Taking the expectation on both sides, we have $\mathbb{E}[X^k e^{sX}] \leq (\frac{k}{(a-s)e})^k M_X(a) < \infty$.

   (c) Let $t = hX \geq 0$, then it is equivalent to show $e^t - 1 \leq te^t$, which is obvious from the inequality $1 - t \leq e^{-t}$.

   (d) From (c), $(e^{hX} - 1)/h$ is bounded by Xe^{hX}, which is an integrable function for sufficiently small $h < a$ from (b). Therefore, we can apply DCT and get the desired result.

3. The existence of the limit follows from L'Hospital's rule.

$$
\begin{aligned}
\lim_{x\to\infty} xe^{x^2/(2\sigma^2)} P(X \geq x) &= xe^{x^2/(2\sigma^2)} \int_x^\infty \frac{1}{\sqrt{2\pi}\sigma} e^{-u^2/(2\sigma^2)} du \\
&= \lim_{x\to\infty} \frac{\int_x^\infty \frac{1}{\sqrt{2\pi}\sigma} e^{-u^2/(2\sigma^2)} du}{e^{-x^2/(2\sigma^2)}/x} \\
&= \lim_{x\to\infty} \frac{-\frac{1}{\sqrt{2\pi}\sigma} e^{-x^2/(2\sigma^2)}}{(-x^2 e^{-x^2/(2\sigma^2)}/\sigma^2 - e^{-x^2/(2\sigma^2)})/x^2} \\
&= \lim_{x\to\infty} \frac{\sigma^2 x^2}{x^2+\sigma^2} \frac{1}{\sqrt{2\pi}\sigma} = \frac{\sigma}{\sqrt{2\pi}}
\end{aligned}
$$

Another method relies on integral by parts:

$$
\begin{aligned}
xe^{x^2/(2\sigma^2)} \int_x^\infty \frac{1}{\sqrt{2\pi}\sigma} e^{-u^2/(2\sigma^2)} du &= \frac{1}{\sqrt{2\pi}\sigma} xe^{x^2/(2\sigma^2)} (\int_x^{+\infty} \frac{\sigma^2}{u} de^{-u^2/(2\sigma^2)}) \\
&= \frac{\sigma}{\sqrt{2\pi}} \left(1 - xe^{x^2/(2\sigma^2)} \int_x^{+\infty} \frac{1}{u^2} e^{-u^2/(2\sigma^2)} du\right)
\end{aligned}
$$

The last time is $o(1)$. Therefore, the limit value is $\frac{\sigma}{\sqrt{2\pi}}$.

4. (i) Let the CDF of X_1 be $F(x)$, PDF of X_1 be $f(x)$. Then

$$\begin{aligned}P(\min(X_1, \dots, X_n) = X_1) &= P(X_i \geq X_1, i = 2, \dots, n) \\ &= \int_{x_i \geq x_1, i=2,\dots,n} f(x_1)f(x_2)\dots f(x_n)dx_1dx_2\dots dx_n \\ &= \int_{\mathbb{R}} f(x_1)dx_1 \left[\int_{x_1}^{+\infty} f(x_2)dx_2\right]^{n-1} \\ &= \int_{\mathbb{R}} f(x_1)(1 - F(x_1))^{n-1}dx_1 \\ &= \int_0^1 (1-u)^{n-1}du = \frac{1}{n}\end{aligned}$$

(ii) For Bernoulli random variables, we have

$$\begin{aligned}P(\min(X_1, \dots, X_n) = X_1) &= P(X_i \geq X_1, i = 2, \dots, n) \\ &= P(X_1 = 1, X_2 = 1, \dots, X_n = 1) + P(X_1 = 0) \\ &= 1 - p + p^n\end{aligned}$$

(iii) When X_i follows exponential distribution parameterized by λ_i, we have $F_i(x) = 1 - \exp(-\lambda_i x)$ and $f_i(x) = -\lambda_i \exp(-\lambda_i x)$.

$$\begin{aligned}P(\min(X_1, \dots, X_n) = X_1) &= P(X_i \geq X_1, i = 2, \dots, n) \\ &= \int_{x_i \geq x_1, i=2,\dots,n} f_1(x_1)f_2(x_2)\dots f_n(x_n)dx_1dx_2\dots dx_n \\ &= \int_{\mathbb{R}} f_1(x_1)dx_1[\prod_{i=2}^{n}\int_{x_1}^{+\infty} f_i(x)dx] \\ &= \int_{\mathbb{R}} f_i(x)\prod_{i=2}^{n}(1 - F_i(x))dx \\ &= \lambda_1\int_0^{+\infty} \exp(-x\sum_{i=1}^{n}\lambda_i)dx = \frac{\lambda_1}{\sum_{i=1}^{n}\lambda_i}\end{aligned}$$

5. We claim that $\beta_0 = \sqrt{2}$. Consider $P(X \leq \beta\sqrt{\log n})$, which equals $P(X_1 \leq \beta\sqrt{\log n})^n$.

$$\begin{aligned}P(X_1 \leq \beta\sqrt{\log n})^n &= \left(\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\beta\sqrt{\log n}} e^{-x^2/2}dx\right)^n \\ &= \exp(n\log\left(1 - \frac{1}{\sqrt{2\pi}}\int_{\beta\sqrt{\log n}}^{+\infty} e^{-x^2/2}dx\right))\end{aligned}$$

From Problem 3, we have $\int_x^\infty e^{-u^2/2}du \sim \frac{\exp(-x^2/2)}{x}$. The integral $\int_{\beta\sqrt{\log n}}^{+\infty} e^{-x^2/2}dx$ can be approximated by $\frac{n^{-\beta^2/2}}{\beta\sqrt{\log n}}$. Therefore, $P(X \leq \beta\sqrt{\log n}) \sim \exp(-n^{1-\beta^2/2}\frac{1}{\sqrt{2\pi}\log n\beta})$. When $\beta \geq \beta_0 = \sqrt{2}$, $P(X \leq \beta\sqrt{\log n}) \to 1$ and $\lim_{n\to\infty} P(X \geq \beta\sqrt{\log n}) = 0$ holds; When $\beta < \beta_0 = \sqrt{2}$, $P(X \leq \beta\sqrt{\log n}) \to 0$ and $\lim_{n\to\infty} P(X \geq \beta\sqrt{\log n}) = 1$ holds.

6. We will derive a recursive formula for $\mathrm{Var}[X_n]$. First we have $\mathbb{E}[X_n] = \mu^n$. Let W_1 follow the offspring distribution. Using the law of total variance, we have $\mathrm{Var}[X_n] = \mathbb{E}[\mathrm{Var}[X_n|X_{n-1}]] + \mathrm{Var}[\mathbb{E}[X_n|X_{n-1}]]$. The random variable $X_n|X_{n-1}$ has mean $X_{n-1}\,\mathbb{E}[W_1]$, variance $X_{n-1}\,\mathrm{Var}[W_1]$. Therefore, $\mathrm{Var}[X_n] = \mathbb{E}[X_{n-1}]\,\mathrm{Var}[W_1] + \mathrm{Var}[X_{n-1}]\,\mathbb{E}[W_1]^2 = \sigma^2\mu^{n-1} + \mu^2\,\mathrm{Var}[X_{n-1}]$, with the initial condition $\mathrm{Var}[X_0] = 0$. As a result,

$$\mathrm{Var}[X_n] = \begin{cases} n\sigma^2 & \mu = 1 \\ \sigma^2\mu^{n-1}\frac{\mu^n-1}{\mu-1} & \mu \neq 1 \end{cases}$$

References

[1] Ye, Min. "Exact recovery and sharp thresholds of Stochastic Ising Block Model." arXiv preprint arXiv:2004.05944 (2020).