解答例

STEP II 1991、第 16 問

確率・統計 ・ 二項分布の正規近似

> **STEP II 1991、問 16**
>
> 先に進む前に、まず問題を読んでおこう。問題は STEP Support Programme の問題データベースにある：
>
> step.maths.org/questions/91-s2-q16

鍵となる考え方

不足量（deficit）、すなわち水位がダムの上端からどれだけ下にあるかを追跡する。放水によって不足量は一定の割合で増え、にわか雨が 1 回降るごとにちょうど V だけ減る。あふれるとは、不足量が負になることである。2 日間の問題では、あふれる可能性のある雨の降り方は 1 通りしかなく、そのとき関係するのは一様分布に従う 1 つの時刻だけである。夏の問題では放水が止められるので、すべてはにわか雨の回数を数えることに帰着し、その回数は二項分布（binomial distribution）に従う。

解答

小問 (i)：2 日間の休暇

休暇が始まる真夜中から日単位で時間を測り、時刻 t における不足量を $D(t)$ とする。すると $D(0) = V$ である。にわか雨とにわか雨の間では D は割合 u で増加し、にわか雨が 1 回降るごとに V だけ減少する。D が一度でも負になれば、ダムはあふれる。（問題の意図どおり、2 日間を通して放水を続けられるだけの水が貯水池にあるものと仮定する。）

1 日目に雨が降る事象を R_1、2 日目に雨が降る事象を R_2 とし、雨の降る時刻（1 日に対する割合で表す）を T_1、T_2 とする。これらはそれぞれ $[0,1]$ 上の一様分布（uniform distribution）に従う。「任意の 1 日について確率 p」という表現に暗に含まれているとおり、各日の出来事は独立（independent）である。

► **雨が降らない場合：**不足量は増えるだけであり、あふれることはない。

- **1 日だけ雨が降る場合**（時刻 s とする）：にわか雨の直後には $D = V + us - V = us \geq 0$ となるので、あふれない（確率 0 の事象 $s = 0$ を除く）。
- **両日とも雨が降る場合：** 1 回目のにわか雨の後は $D = uT_1 \geq 0$ であり、安全である。2 回目のにわか雨の直後、すなわち時刻 $1 + T_2$ では

$$D = V + u(1 + T_2) - 2V = u(1 + T_2) - V.$$

  T_1 が消えていることに注意しよう。1 回目のにわか雨がいつ降っても、放水は同じ割合で続くからである。

したがって、あふれるのは両日とも雨が降り、かつ $u(1 + T_2) < V$、すなわち $T_2 < \frac{V}{u} - 1$ となるときであり、そのときに限る。$\frac{1}{2}V \leq u \leq V$ ならば数 $\frac{V}{u} - 1$ は $[0, 1]$ に含まれ、T_2 は一様分布に従うので

$$\mathrm{P}(\text{あふれる}) = p^2 \left(\frac{V}{u} - 1 \right).$$

これを Q に等しいとおくと

$$\frac{V}{u} = 1 + \frac{Q}{p^2} \quad \Longrightarrow \quad u = \frac{p^2 V}{p^2 + Q}.$$

条件 $Q < p^2$ があるからこそ、この結果は矛盾なく成り立つ。$0 < Q/p^2 < 1$ から、仮定どおり $\frac{1}{2}V < u < V$ が得られるのである。（この条件がなければ解は存在しない。雨の日が 2 日必要なので、あふれる確率は決して p^2 を超えないからである。）

小問 (i)：夏の休暇

今度は $u = 0$ なので、不足量は決して増えない。不足量は kV から始まり、にわか雨のたびに V ずつ減る。よって 18 日のうち雨の降る日数を N とすると

$$\text{あふれる} \iff NV > kV \iff N > k.$$

各日は独立で、それぞれ確率 $\frac{1}{3}$ で雨が降るので、$N \sim \mathrm{B}(18, \frac{1}{3})$ であり、平均（mean）は 6、分散（variance）は $18 \cdot \frac{1}{3} \cdot \frac{2}{3} = 4$ である。

> ? **考えてみよう** ここで「適切な」近似とはどれか。また、なぜ連続性補正が必要なのか。

$np = 6$、$n(1 - p) = 12$ なので、正規近似（normal approximation）$N \approx \mathrm{N}(6, 4)$ は妥当である。（p が小さくないので、ポアソン（Poisson）近似は適切でない。）N は整数値をとるので、整数 k に対して事象 $N > k$ は事象 $N \geq k + 1$ と同じであり、これを $X \sim \mathrm{N}(6, 4)$ として $X > k + \frac{1}{2}$ で近似する。これを連続性補正（continuity correction）

という。必要な条件は

$$\mathrm{P}\left(Z > \frac{k+\frac{1}{2}-6}{2}\right) < \frac{1}{10} \iff \frac{k-5.5}{2} > 1.2816 \iff k > 8.06.$$

具体的には、$k = 8$ のとき $\mathrm{P}(Z > 1.25) \approx 0.106$ となり大きすぎるが、$k = 9$ のとき $\mathrm{P}(Z > 1.75) \approx 0.040$ となり条件を満たす。あふれるかどうかは $N > k$ であるかどうかだけで決まるので、8 と 9 の間の整数でない k を選んでも $k = 9$ より良くはならない。したがって技師は、貯水池の水位を

$$k = 9,$$

すなわち上端から $9V$ 下まで下げておくべきである。

小問 (ii)

今度は独立な半日が 36 個あり、それぞれ確率 $\frac{1}{6}$ で雨が降る。放水は依然として止まっているので、あふれるのはやはり $N > k$ のときであり、今度は $N \sim \mathrm{B}(36, \frac{1}{6})$ である。平均は依然として 6 だが、分散は $36 \cdot \frac{1}{6} \cdot \frac{5}{6} = 5$ であり、4 より大きい。

平均が同じで分散が大きければ、上側の裾が重くなる。すなわち、平均より上の各 k に対して、$\mathrm{P}(N > k)$ は以前より大きくなる。よって必要な k は**大きく**なる。連続性補正を用いて $\mathrm{N}(6, 5)$ で近似すると、必要な条件は

$$\frac{k-5.5}{\sqrt{5}} > 1.2816 \iff k > 8.37,$$

であり、以前の $k > 8.06$ と比べて大きい。（ここでは正規近似により $\mathrm{P}(N \geq 9) \approx 0.132$ となり、小問 (i) の 0.106 より大きい。）このように、しきい値は上がる。k を整数に切り上げれば、どちらの場合も $k = 9$ となる。しかし $k = 9$ における裾の確率は、0.04 ではなく約 0.06 になるので、安全の余裕は小さくなる。

検算

- $p = 0.5$、$Q = 0.1$、$u = \frac{0.25}{0.35}V \approx 0.714V$ として 2 日間の休暇のモンテカルロ・シミュレーション（Monte Carlo simulation）を行ったところ（400,000 回）、あふれる頻度は 0.0996 であった。p と Q を他の値にしても、同様によく一致した。
- 極限の場合：$Q \to p^2$ のとき $u \to \frac{1}{2}V$ となる。これは、雨の日が 2 日あればまだあふれうる最も遅い放水の割合である。$Q \to 0$ のとき $u \to V$ となり、この割合では 2 回目のにわか雨で決してあふれない。
- 二項分布の裾の正確な値によって、正規近似が確かめられる。$\mathrm{B}(18, \frac{1}{3})$ では $\mathrm{P}(N \geq 9) = 0.1076$、$\mathrm{P}(N \geq 10) = 0.0433$ である。$\mathrm{B}(36, \frac{1}{6})$ では $\mathrm{P}(N \geq 9) = 0.1332$、$\mathrm{P}(N \geq 10) = 0.0652$ である。

よくある間違い

- 2 日間の問題で、1 回目の雨の時刻 T_1 が関係すると考えてしまう。放水の割合は一定なので、これは打ち消し合う。
- 連続性補正を忘れる。補正をしないと不等式は $k > 8.56$ となり、N が整数であるという事実が見えなくなる。
- 平均が等しいからという理由で、(ii) の答えは「同じ」と言ってしまう。裾を支配するのは散らばりである。

STEP のその先へ

雨がランダムな瞬間に降り、水が一定の割合で流れ出ていくという設定は、応用確率論における古典的なモデルであり、Moran がダムのために発展させた貯水理論（storage theory）と呼ばれる。これは待ち行列（queue）や保険の破産（ruin）の理論と密接に関係している。小問 (ii) において、平均を固定したまま 1 日をより多くのより小さな区間に分割していくと、$\mathrm{B}(n, 6/n)$ は平均 6、分散 6 のポアソン分布に近づき、雨の到来はポアソン過程（Poisson process）になる。さらに学ぶには、Grimmett and Welsh, *Probability: An Introduction*（確率論入門）を参照するとよい。

Beyond STEP 解答集：https://beyondstep.keithmcnulty.org/ja/。解答本文は CC0 1.0 である。問題そのものは © Cambridge University Press & Assessment に帰属し、ここには掲載していない。英語版から機械翻訳したものであり、まだ校閲を経ていない。