解答例

STEP II 1993、第 16 問

確率・統計 ・ ポアソン過程

> **STEP II 1993、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/93-s2-q16

鍵となる考え方

この問題には独立な 2 つの乱雑さの源がある。1 つはバスであり、これは幾何分布（geometric distribution）に従う待ち時間を与える。もう 1 つは乗客であり、これはポアソン過程（Poisson process）をなす。最初のバスがいつ来るかで条件付けをする（condition）。それがわかれば、待っている人数は、長さのわかった時間区間におけるポアソン分布に従う個数にすぎない。小問 (iii) では全確率の公式（law of total probability）を用いる。本当の問題は、その結果得られる無限級数を、数表と電卓だけで有効数字 2 桁まで求めるにはどうすればよいかである。

解答

時間は 8:00 からの経過分数で測る。最初のバスが到着する時刻を T とし、区間 $(0, t]$ に到着する乗客の人数を X_t とする。乗客の到着は 1 分あたり率 4 のポアソン過程であるから、$X_t \sim \mathrm{Po}(4t)$ である。バスが来うるのは $t = 1, 2, 3, \ldots$ のときだけであり（「8 時を少し過ぎたところ」とは、8:00 に来うるバスはすでに来て去った後であることを意味する）、バスは乗客とは独立である。

小問 (i)

最初のバスが 8 時 n 分に到着するのは、ちょうど $1, 2, \ldots, n-1$ 分にはバスが来ず、かつ n 分にはバスが来るときである。これらの事象は独立であるから、

$$\mathrm{P}(T = n) = \left(\tfrac{3}{4}\right)^{n-1} \tfrac{1}{4}, \qquad n = 1, 2, 3, \ldots$$

したがって T はパラメータ $\frac{1}{4}$ の幾何分布に従う。（確認として、これらの確率の和は $\frac{1}{4}/(1-\frac{3}{4}) = 1$ である。）

小問 (ii)

8:05 に待っている人々は、ちょうど $(0,5]$ に到着した人々である。8:05 より前にバスは来ていないので、誰も去っていない。よって、$T = 5$ という条件の下での X_5 の分布を求めればよい。

> **考えてみよう** 最初のバスが 8:05 に来たと知ることで、乗客の人数の分布は変わるだろうか？

変わらない。事象 $T = 5$ はバスだけに依存し、バスは乗客と独立である。よって $T = 5$ で条件付けても X_5 の分布は変わらず、$X_5 \sim \mathrm{Po}(4 \times 5) = \mathrm{Po}(20)$ である。したがって

$$\mathrm{P}(m \text{ 人が待っている} \mid T = 5) = \frac{\mathrm{e}^{-20}\, 20^m}{m!}, \qquad m = 0, 1, 2, \ldots$$

小問 (iii)

バスが全員を乗せられないのは、ちょうど 25 人より多くの人が待っているとき、すなわち $X_T \geq 26$ のときである。この事象を F とする。T で条件付けし、(ii) と同様に独立性を用いると、

$$\mathrm{P}(F) = \sum_{n=1}^{\infty} \mathrm{P}(T = n)\, \mathrm{P}(X_n \geq 26) = \sum_{n=1}^{\infty} \left(\tfrac{3}{4}\right)^{n-1} \tfrac{1}{4}\, \mathrm{P}(\mathrm{Po}(4n) \geq 26).$$

ここで、この無限級数を有限の手間の計算に変え、しかも誤差を 0.005 よりも十分小さく抑える方針が必要である。求めたいのは有効数字 2 桁であり、答えは 0.1 程度の大きさになることがわかる。

ステップ 1：余事象を扱い、裾の部分を容易にする。 $w_n = \left(\frac{3}{4}\right)^{n-1} \frac{1}{4}$、$q_n = \mathrm{P}(\mathrm{Po}(4n) \leq 25)$ とおく。$\sum w_n = 1$ であるから、

$$\mathrm{P}(F) = 1 - \sum_{n=1}^{\infty} w_n\, q_n.$$

この形のほうがよいのは、n が大きいとき q_n がきわめて小さいからである。$n \geq 11$ では平均 $4n \geq 44$ は 25 を大きく上回る。n が増えると q_n は減少する（時間が長いほど乗客は多い）から、

$$\sum_{n \geq 11} w_n q_n \leq q_{11} \sum_{n \geq 11} w_n = q_{11} \left(\tfrac{3}{4}\right)^{10}.$$

正規近似（normal approximation）により $q_{11} \approx \Phi((25.5-44)/\sqrt{44}) = \Phi(-2.79) \approx 0.003$ であるから、この裾の部分は 0.0002 未満であり、無視してよい。残るのは 10 項である。

ステップ 2：最初のいくつかの項。 $n=1$ と $n=2$（平均 4 と 8）では、26 人以上の乗客が来ることはきわめて起こりにくいので、必要な精度をはるかに超えて $q_1 = q_2 = 1$ としてよい。よってこれらの寄与は $w_1 + w_2 = \frac{1}{4} + \frac{3}{16}$ である。

ステップ 3：中間の項は数表から。 $n = 3,4,5$（平均 12、16、20）については、$q_n = \mathrm{P}(\mathrm{Po}(4n) \le 25)$ をポアソン分布の累積分布表から直接読み取る。この表は通常、平均 20 までを扱っている。表に平均 24 と 28 も載っていれば、$n = 6,7$ にもそれを使う。

ステップ 4：大きい平均は正規近似で。 残りの n から 10 まで（平均は最大 40）については、連続修正（continuity correction）を施した $\mathrm{Po}(\mu) \approx \mathrm{N}(\mu,\mu)$ を用いる：

$$q_n \approx \Phi\left(\frac{25.5 - 4n}{\sqrt{4n}}\right),$$

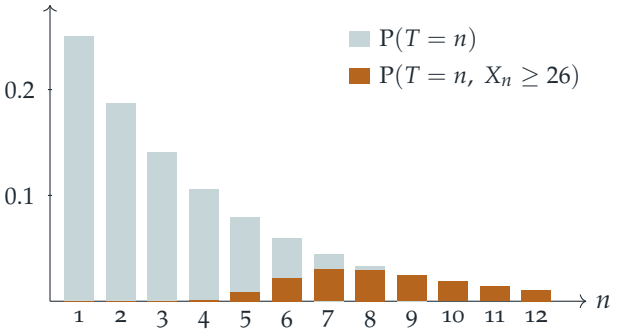
ここで Φ は正規分布表から読み取る。これらの平均は近似がよく効くほど十分大きい。さらに重要なのは、重み w_n がすでに小さい（$w_8 \approx 0.033$、$w_{10} \approx 0.019$）ので、q_n に千分の数程度の誤差があっても $\mathrm{P}(F)$ の変化は 0.001 未満にとどまることである。

ステップ 5：まとめる。 電卓で $w_1, \ldots, w_{10}$ を計算し、

$$\mathrm{P}(F) \approx 1 - \sum_{n=1}^{10} w_n q_n,$$

を求めてから、有効数字 2 桁に丸める。無視した部分や近似した部分はそれぞれ高々 0.001 程度であるから、合計が丸めの境界にきわめて近くならない限り、丸めは安全である。もし近くなったとしても、この方法ならどの項を精密にすればよいかがはっきりわかる。

問題文はこの計算を実行することまでは求めていないが、実行すると $\mathrm{P}(F) \approx 0.19$ となる。下のグラフはこの確率がどこから来るかを示している。その大部分は最初のバスが 8:06 から 8:10 に来る場合から生じる。このとき、待ち行列の人数の期待値は 24 から 40 であり、定員に近いか定員を超えている。



検算

- (i) では、確率の和は 1 であり、$\mathrm{P}(T=1)=\frac{1}{4}$ となっていて、あるべき通りである。
- (ii) では、待っている人数の平均は 20 であり、これは 1 分あたり 4 人が 5 分間来ることに対応する。
- (iii) では、厳密な和（計算機で計算したもの）は 0.1904 である。2×10^5 回の朝をモンテカルロ・シミュレーション（Monte Carlo simulation）すると 0.189 が得られる。上の数表と正規近似による方法では 0.1912 となる。3 つとも丸めると 0.19 になる。

よくある間違い

- 境界を 1 つずらしてしまうこと。バスが全員を乗せられないのは 26 人以上が待っているときだけであるから、求める確率は $1-\mathrm{P}(\mathrm{Po}(4n)\leq 25)$ であって、$1-\mathrm{P}(\mathrm{Po}(4n)\leq 24)$ ではない。
- 待ち行列の人数を $\mathrm{Po}(4\,\mathrm{E}[T])=\mathrm{Po}(16)$ として扱うこと。待ち時間は確率的であるから、平均時間を代入するのではなく、ポアソン確率を T について平均しなければならない。
- 級数を打ち切ってよい理由を述べずに、無限級数のまま放置すること。無視した裾の部分が小さすぎて有効数字 2 桁目に影響しないことを示さなければならない。

STEP のその先へ

待っている人数 X_T は、平均が確率的に決まるポアソン分布に従う個数、すなわちポアソン分布の**混合**（mixture）である。T が幾何分布に従うので、X_T は簡潔な閉じた形の母関数（generating function）をもつことがわかる。また、このような条件付けに

よる議論は、待ち行列理論（queueing theory）の基礎であり、確率論の最初の講義で出会う複合分布や混合分布の基礎でもある。級数をどこで打ち切り、誤差をどう評価するかという問題は、数値解析（numerical analysis）の出発点である。Grimmett and Welsh, *Probability: An Introduction*（確率論入門）は、ポアソン過程と母関数をちょうどよい水準で扱っている。

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。